OM nucleic - nucleic search, using sw model

Run on: April 30, 2004, 14:44:42 ; Search time 3927 Seconds
(without alignments)
9966.580 Million cell updates/sec

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Title: US-10-202-687-1
Perfect score: 903
Sequence: 1 atggacctgcccccgcagct.........ggggcaagtcccagaagtaa 903
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 3470272 seqs, 2167151995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	903	100.0	903	6	AR308734	AR308734 Sequence
2	903	100.0	903	6	AX549189	AX549189 Sequence
3	903	100.0	903	6	AX675043	AX675043 Sequence
4	903	100.0	923	9	AF024687	AF024687 Homo sapi
5	903	100.0	34791	9	HSU62631	U62631 Human B-cel
6	898.2	99.5	903	6	AR308751	AR308751 Sequence
7	843.8	93.4	903	9	AB095743	AB095743 Macaca fa
8	614	68.0	903	10	AB095746	AB095746 Mesocrice
9	604.4	66.9	903	6	AX675045	AX675045 Sequence
10	604.4	66.9	903	10	AB095745	AB095745 Mus muscu
11	602.8	66.8	1819	10	AF539809	AF539809 Mus muscu
c 12	602.8	66.8	209873	2	AC087143	AC087143 Mus muscu
13	598	66.2	154673	2	AC079472	AC079472 Mus muscu
14	590	65.3	903	10	AB095744	AB095744 Rattus no
15	590	65.3	907	10	AF539810	AF539810 Rattus no
16	588.4	65.2	238725	2	AC115260	AC115260 Rattus no
17	588.4	65.2	253523	2	AC126062	AC126062 Rattus no
c 18	582	64.5	154673	2	AC079472	AC079472 Mus muscu
19	134.6	14.9	207989	2	AC146547	AC146547 Gasterost
20	123.4	13.7	1841	6	AR140519	AR140519 Sequence
21	123.4	13.7	1841	6	E41325	E41325 cDNA clone
22	121.8	13.5	1041	6	AR308737	AR308737 Sequence
23	121.8	13.5	1041	6	AX224754	AX224754 Sequence
24	121.8	13.5	1041	6	AX224756	AX224756 Sequence
25	121.8	13.5	1041	6	AX549319	AX549319 Sequence
26	121.8	13.5	1061	9	AF024688	AF024688 Homo sapi
27	121.8	13.5	1061	9	AF024689	AF024689 Homo sapi
28	121.8	13.5	1694	9	BC035657	BC035657 Homo sapi
29	121.8	13.5	98713	9	AC002511	AC002511 Human DNA
30	118.6	13.1	1041	6	AR308752	AR308752 Sequence
31	118.6	13.1	1158	9	AY431102	AY431102 Homo sapi
32	84.6	9.4	1534	6	BD137966	BD137966 Protease-
33	84.6	9.4	1534	9	AF080214	AF080214 Homo sapi
34	84.6	9.4	4895	6	AR108643	AR108643 Sequence
35	84.6	9.4	4895	6	AR223400	AR223400 Sequence
36	84.6	9.4	4895	9	AX242695	AX242695 Sequence
37	84.6	9.4	4895	9	AX549230	AX549230 Sequence
38	84.6	9.4	4895	9	AF055917	AF055917 Homo sapi
39	84.6	9.4	4925	6	AX211769	AX211769 Sequence
40	84.6	9.4	11828	9	AF384819	AF384819 Homo sapi
41	84.6	9.4	121886	2	BX663530	BX663530 Gallus ga
42	84.6	9.4	248281	9	AC008737	AC008737 Homo sapi
43	84.4	9.3	131743	2	BX663529	BX663529 Gallus ga
c 44	83.4	9.2	84653	2	BX663523	BX663523 Gallus ga
c 45	83.4	9.2	252295	2	BX663526	BX663526 Gallus ga

ALIGNMENTS

RESULT 1
AR308734

LOCUS   AR308734   903 bp   DNA   linear   PAT 12-JUN-2003
DEFINITION Sequence 249 from patent US 6655339.
ACCESSION   AR308734
VERSION   AR308734.1 GI:31170263
KEYWORDS   .
SOURCE   Unknown.
ORGANISM   Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 903)
AUTHORS   Liaw,C.W., Behan,D.P. and Chalmers,D.T.
TITLE   Non-endogenous, constitutively activated human protein-coupled
            receptors
JOURNAL   Patent: US 6655339-A 249 29-APR-2003;

Pred. No. is the number of results predicted by chance to have a

```
FEATURES             Location/Qualifiers
     source          1..903
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Query Match        100.0%;   Score 903;  DB 6;  Length 903;
Best Local Similarity 100.0%;  Pred. No. 1.5e-132;
Matches  903;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      1 ATGGACCTGCCCCGCAGCTCTCCTTCGGCCTCTATGTGGCCGCTGGGCTTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGGACCTGCCCCGCAGCTCTCCTTCGGCCTCTATGTGGCCGCTGGGCTTC 60

Qy     61 CGGCTCAACGTCCTGGCCATCGAGGCGCGACGGCCCAGCGCCTCCGTCCGTCTCACCCCT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CGGCTCAACGTCCTGGCCATCGAGGCGCGACGGCCCAGCGCCTCCGTCCGTCTCACCCCT 120

Qy    121 AGCCTGGTCTACGCCCTGAACCTGGGCTGTCTCCGACCTGCTGCTGACAGTCTCTCTGCCC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AGCCTGGTCTACGCCCTGAACCTGGGCTGTCTCCGACCTGCTGCTGACAGTCTCTCTGCCC 180

Qy    181 CTGAAGGGCGTGGAGGGCGCTAGCCTCCGGGGGCTTCCTGGGGGCTTCCTGCGGTGCCCC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CTGAAGGGCGTGGAGGGCGCTAGCCTCCGGGGGCTTCCTGGGGGCTTCCTGCGGTGCCCC 240

Qy    241 GTCTTCGGCGGTGGCCACTTCTTCCCACTTCATGCGGCGGGCGGGGCTTCCTGGCCGCCCTG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GTCTTCGGCGGTGGCCACTTCTTCCCACTTCATGCGGCGGGCGGGGCTTCCTGGCCGCCCTG 300

Qy    301 AGTGCAGGCCGCTACCTGGGCAGCAGCCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGTGCAGGCCGCTACCTGGGCAGCAGCCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCG 360

Qy    361 TGCTATTCCTGGGGGGGTGTGCGCGGCCATCTGGGCGCCTCGTCCTGTCACCTGGGGTCTG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGCTATTCCTGGGGGGGTGTGCGCGGCCATCTGGGCGCCTCGTCCTGTCACCTGGGGTCTG 420

Qy    421 GTCTTTGGGTTGAGGGCTCCAGGAGGCTGCGCTGAGCGGTGTGCTGACCACAGCAACACCTCCCTGGGCATC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GTCTTTGGGTTGAGGGCTCCAGGAGGCTGCGCTGAGCGGTGTGCTGACCACAGCAACACCTCCCTGGGCATC 480

Qy    481 AACACACCGGTCAACGGCTCTCCCGTCTCCCGGAGGCCTGGAGCCCGGACCCCGGCCTCTGCCCGGC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AACACACCGGTCAACGGCTCTCCCGTCTCCCGGAGGCCTGGAGCCCGGACCCCGGCCTCTGCCCGGC 540

Qy    541 CCGGCCCGCTTCAGCCCTTGGCCCAGCTTCCTGTACCCCAATCTAGGAGGCTTCCTGGGGGCGGAAGCTG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CCGGCCCGCTTCAGCCCTTGGCCCAGCTTCCTGTACCCCAATCTAGGAGGCTTCCTGGGGGCGGAAGCTG 600

Qy    601 TGCTACGTGGGCTGCTCCGGGCACTGGCACTGGTGCTGCTGCTCTTCCGCGCTGCGGGCCCCTCCTCACGCTGTCTGCGCTAGGAGACCCTAC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TGCTACGTGGGCTGCTCCGGGCACTGGCACTGGTGCTGCTGCTCTTCCGCGCTGCGGGCCCCTCCTCACGCTGTCTGCGCTAGGAGACCCTAC 720

Qy    721 AACGGCTCCAAACGTGGCCAGCTTCCTGTACCCCAATCTAGGAGGCTTCCTGGGGGCGGAAGCTG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 AACGGCTCCAAACGTGGCCAGCTTCCTGTACCCCAATCTAGGAGGCTTCCTGGGGGCGGAAGCTG 780

Qy    781 GGGCTCATCACGGGTGCCTGGAGTGTGGTGCTTAATCCGGTTACTTGGGA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GGGCTCATCACGGGTGCCTGGAGTGTGGTGCTTAATCCGGTTACTTGGGA 840

Qy    841 AGGGGTCCTGGCCTGAAGACAGTGTGCGGCAAGAACGCAAGGGGGCAAGTCCCAGAAG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 AGGGGTCCTGGCCTGAAGACAGTGTGCGGCAAGAACGCAAGGGGGCAAGTCCCAGAAG 900

Qy    901 TAA 903
          |||
Db    901 TAA 903
```

```
Db   121 AGCCTGGTCTACGCCCTGAACCTGGGCTGCTCCGACGCACAGGCGGAAGCTG 180
                                                              180
Qy   181 CTGAAGGCGGTGGAGGCGCTAGGCCGCTGCTGCTCGGAGGACCCTAC 240
             |||||||||||||||||||||||||||||||||||||||||||||
Db   181 CTGAAGGCGGTGGAGGCGCTAGGCCGCTGCTGCTCGGAGGACCCTAC 240
Qy   241 GTCTTCGGCGTGGCGCCACTTCTTCCACTCTATGCGGCGGGGCTTCCTGGCCGCCCTG 300
             |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GTCTTCGGCGTGGCGCCACTTCTTCCACTCTATGCGGCGGGGCTTCCTGGCCGCCCTG 300
Qy   301 AGTGCAGGCGCTACCTGGGAGCAGCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCG 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 AGTGCAGGCGCTACCTGGGAGCAGCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCG 360
Qy   361 TGCTATTCCTGGGGGGTGTGCGGCGGCCATCGTGCCTCGTCCCTGTGTCACCTGGGTCTG 420
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TGCTATTCCTGGGGGGTGTGCGGCGGCCATCGTGCCTCGTCCCTGTGTCACCTGGGTCTG 420
Qy   421 GTCTTTGGGTTGGAGGCTCCAGGAGGCTGGACCGGCCTTGGGACCCGGCCTCTGCGCATC 480
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GTCTTTGGGTTGGAGGCTCCAGGAGGCTGGACCGGCCTTGGGACCCGGCCTCTGCGCATC 480
Qy   481 AACACACCGGTCAACGGCTCTCCGGTCTCGGAGGCCTGGAGACCCGGCCTCTGCGCGGC 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 AACACACCGGTCAACGGCTCTCCGGTCTCGGAGGCCTGGAGACCCGGCCTCTGCGCGGC 540
Qy   541 CCGGCCCGCTTCAGCCTCTTCAGCCTCTCCCGGCCTCCGGCCTGACGGCGCACAGGCGGAAGCTG 600
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 CCGGCCCGCTTCAGCCTCTTCAGCCTCTCCCGGCCTCCGGCCTGACGGCGCACAGGCGGAAGCTG 600
Qy   601 TGCTACGTGGCGCTGCTCCGGCCCGGCCTCCGGCCTGACGGCGCGGAAGCTG 660
             |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 TGCTACGTGGCGCTGCTCCGGCCCGGCCTCCGGCCTGACGGCGCGGAAGCTG 660
Qy   661 CGGGCCGCCTGGGTGGCCGGGCCCCTCCTCACGCTGTGCTCTCGCGTAGGACCCTAC 720
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 CGGGCCGCCTGGGTGGCCGGGCCCCTCCTCACGCTGTGCTCTCGCGTAGGACCCTAC 720
Qy   721 AACGCCTCCAACGTGGCCAGCTTCCTGTACCCCAATCTAGGAGGCTCCTGGCCGGAAGCTG 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 AACGCCTCCAACGTGGCCAGCTTCCTGTACCCCAATCTAGGAGGCTCCTGGCCGGAAGCTG 780
Qy   781 GGGCTCATCACGGGTGCCTGGAGTGGTGCTTAATCCGGCTGGTGACCGGTTACTTGGGA 840
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 GGGCTCATCACGGGTGCCTGGAGTGGTGCTTAATCCGGCTGGTGACCGGTTACTTGGGA 840
Qy   841 AGGGGTCCTGGCCTGAAGACAGTGTGCGGCAAGAACGCAAGGGGCAAGTCCCAGAAG 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 AGGGGTCCTGGCCTGAAGACAGTGTGCGGCAAGAACGCAAGGGGCAAGTCCCAGAAG 900
Qy   901 TAA 903
             |||
Db   901 TAA 903
```

RESULT 4
AF024687
LOCUS   AF024687   923 bp   DNA   linear   PRI 21-NOV-1997
DEFINITION   Homo sapiens putative G protein-coupled receptor (GPR40) gene,
             complete cds.
ACCESSION   AF024687
VERSION   AF024687.1   GI:2612945
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 923)
  AUTHORS   Sawzdargo,M., George,S.R., Nguyen,T., Xu,S., Kolakowski,L.F. and
             O'Dowd,B.F.
  TITLE   A cluster of four novel human G protein-coupled receptor genes
             occurring in close proximity to CD22 gene on chromosome 19q13.1
  JOURNAL   Biochem. Biophys. Res. Commun. 239 (2), 543-547 (1997)

```
Db   601 TGCTACGTGGCGCTGCTCCGGGCCACTGGCCCGGCCTCCGGCCCGCTGCCCGGCGGAAGCTG 660
                                                                          660
Qy   661 CGGGCCGCCTGGGTGGCCGGGCCCCTCCTCACGCTGTGCTCTCGCGTAGGACCCTAC 720
             ||||||||||||||||||||||||||||||||||||||||||||||
Db   661 CGGGCCGCCTGGGTGGCCGGGCCCCTCCTCACGCTGTGCTCTCGCGTAGGACCCTAC 720
Qy   721 AACGCCTCCAACGTGGCCAGCTTCCTGTACCCCAATCTAGGAGGCTCCTGGCCGGAAGCTG 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 AACGCCTCCAACGTGGCCAGCTTCCTGTACCCCAATCTAGGAGGCTCCTGGCCGGAAGCTG 780
Qy   781 GGGCTCATCACGGGTGCCTGGAGTGGTGCTTAATCCGGCTGGTGACCGGTTACTTGGGA 840
             |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 GGGCTCATCACGGGTGCCTGGAGTGGTGCTTAATCCGGCTGGTGACCGGTTACTTGGGA 840
Qy   841 AGGGGTCCTGGCCTGAAGACAGTGTGCGGCAAGAACGCAAGGGGCAAGTCCCAGAAG 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 AGGGGTCCTGGCCTGAAGACAGTGTGCGGCAAGAACGCAAGGGGCAAGTCCCAGAAG 900
Qy   901 TAA 903
             |||
Db   901 TAA 903
```

RESULT 3
AX675043
LOCUS   AX675043   903 bp   DNA   linear   PAT 27-MAR-2003
DEFINITION   Sequence 1 from Patent WO02057783.
ACCESSION   AX675043
VERSION   AX675043.1   GI:29333351
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Andrews,J.L., Briscoe,C.P., Ignar,D.M., Muir,A.I., Sauls,H.R. and
             Tadayyon,M.
  TITLE   Method of screening for gpr40 ligands
  JOURNAL   Patent: WO 02057783-A 1 25-JUL-2002;
             Glaxo Group Limited (GB) ; Smithkline Beecham PLC (GB)
FEATURES   Location/Qualifiers
     source   1..903
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
     CDS   1..903
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD83804.1"
                     /db_xref="GI:29333352"
                     /db_xref="REMTREMBL:CAD83804"
                     /translation="MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVY
                     ALNLGCSDLLLTVSLPLKAVERALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSA
                     GRYLGAAFPLGYQAFRRPCYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGI
                     NTPVNGSPVCLEAWDPASAGPARFSLSLLLFPLPLAITAFCYVGCLRALARSGLITHRR
                     KLRAAWVAGGALLTLLLCVGPYNASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVT
                     GYLGRGPGLKTVCAARTQGGKSQK"
ORIGIN
```

Query Match     100.0%;    Score 903;    DB 6;    Length 903;
Best Local Similarity  100.0%;    Pred. No. 1.5e-132;
Matches  903;    Conservative    0;    Mismatches    0;    Indels    0;    Gaps    0;

```
Qy     1 ATGGACCTGCCCCCGCAGCTCTCCTTCGGCCTCTATGTGGCGGCCTTTGCGCTGGGCTTC 60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 ATGGACCTGCCCCCGCAGCTCTCCTTCGGCCTCTATGTGGCGGCCTTTGCGCTGGGCTTC 60
Qy    61 CCGCTCAACGTCCTGGCCATCCGAGGCGGCGACGGCCCCACGGCCCACGCCCCGGCTCGGTCCTCCACCCCT 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 CCGCTCAACGTCCTGGCCATCCGAGGCGGCGACGGCCCCACGGCCCACGCCCCGGCTCGGTCCTCCACCCCT 120
Qy   121 AGCCTGGTCTACGCCCTGAACCTGGGCTGCTCCGACCTGCTGCTGCTCTCTGCCC 180
             |||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
MEDLINE   98008875
PUBMED    9344866
REFERENCE 2  (bases 1 to 923)
AUTHORS   O'Dowd,B.F.
TITLE     Direct Submission
JOURNAL   Submitted (15-SEP-1997) Department of Pharmacology, University of
          Toronto, 8 Taddle Creek Rd., Toronto, ON M5S 1A8, Canada
FEATURES             Location/Qualifiers
     source          1..923
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /map="19q13.1"
     gene            11..913
                     /gene="GPR40"
     CDS             11..913
                     /gene="GPR40"
                     /codon_start=1
                     /product="putative G protein-coupled receptor"
                     /protein_id="AAB86710.1"
                     /db_xref="GI:2612946"
                     /translation="MDLPPQLSFGLVVAAFALGFPLNVLAIRGATAHARLLTPSLVY
                     ALNLGCSDLLLTVSLPLKAVERALASGAWPLPASLCPVFAVAHFFPLYAGGFLAALSA
                     GRYLGAAFPLGYQAFRRPCYSWGVCAAIWALVCHLGLVFGLEAPGGWLDHSNTSLGI
                     NTPVNGSPVCLEAWDPASAGPARPSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRR
                     KLRAAMWVAGGALLTLLLCVGPYNASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVT
                     GYLGRGPGLKTVCAARTQGGKSQK"

ORIGIN

Query Match      100.0%;  Score 903;  DB 9;  Length 923;
Best Local Similarity   100.0%;
Matches  903;    Conservative    0;    Mismatches    0;    Indels    0;    Gaps    0;
```

```
Qy      1 ATGGACCTGCCCCCGCAGCTCTCTTCGGCCTCATGTGGCCGCCTTGCGCTGGGCTTC   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11 ATGGACCTGCCCCCGCAGCTCTCTTCGGCCTCATGTGGCCGCCTTGCGCTGGGCTTC   70

Qy     61 CCGCTCAACGTCCTGGCCATCAGAGGGGCGACAGGCCCACGCGCCACGGCTCACACCCCT  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     71 CCGCTCAACGTCCTGGCCATCAGAGGGGCGACAGGCCCACGCGCCACGGCTCACACCCCT  130

Qy    121 AGCCTGGTCTACGGCCCTGAACCTGGGCTGTCTCCGACCTGCTGACAGTCTCTCTGCCC  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    131 AGCCTGGTCTACGGCCCTGAACCTGGGCTGTCTCCGACCTGCTGACAGTCTCTCTGCCC  190

Qy    181 CTGAAGGCGGTGAGGGCGCTAGCCTCCGGGGCCTGGCCTTCGCCTGTCGCCGTGCCCC  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    191 CTGAAGGCGGTGAGGGCGCTAGCCTCCGGGGCCTGGCCTTCGCCTGTCGCCGTGCCCC  250

Qy    241 GTCTTCGCGGTGGCCCACTTCTTCCCACTCTATGCCGGGCTACCAAGCCTTCCGAGGCCG  300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    251 GTCTTCGCGGTGGCCCACTTCTTCCCACTCTATGCCGGGCTACCAAGCCTTCCGAGGCCG  310

Qy    301 AGTGCAGGCCGCTACCTGGGAGCAGCCTTCCCCCTTGGGCTACCAAGCCTTCCGAGGCCG  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    311 AGTGCAGGCCGCTACCTGGGAGCAGCCTTCCCCCTTGGGCTACCAAGCCTTCCGAGGCCG  370

Qy    361 TGCTATTCCTGGGGGGTGTGCGCGGCGCCATCTGGGCGCCCTGTCACCTGGGTCTG  420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    371 TGCTATTCCTGGGGGGTGTGCGCGGCGCCATCTGGGCGCCCTGTCACCTGGGTCTG  430

Qy    421 GTCTTGGGTTGGGTTGGAGGCTCCAGAGGGCTGCTGACGAGGCTGCTGGCTGGGCATC  480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    431 GTCTTGGGTTGGGTTGGAGGCTCCAGAGGGCTGCTGACGAGGCTGCTGGCTGGGCATC  490

Qy    481 AACACACCGTCAACGTCCTGGAGGCCTTGGACCCGGACCCTCTGCCCGC  540
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    491 AACACACCGTCAACGTCCTGGAGGCCTTGGACCCGGACCCTCTGCCCGC  550

Qy    541 CCGGCCCGCTTCAGCCTCTCTCTGCTCTTTTTCTGCCCGTCTTTTTTCTGCCCATCACAGCCTTC  600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    551 CCGGCCCGCTTCAGCCTCTCTCTGCTCTTTTTCTGCCCGTCTTTTTTCTGCCCATCACAGCCTTC  610
```

```
Qy    601 TGCTACGTGGGCTGCCTCCGGGCACTGGCCCGTCCGCTCGGGCCTGAGCGCCAGGCGGAAGCTG  660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    611 TGCTACGTGGGCTGCCTCCGGGCACTGGCCCGTCCGCTCGGGCCTGAGCGCCAGGCGGAAGCTG  670

Qy    661 CGGGCCGCCTGGGTGGTGGCCGGGCGGGGCCCGGCCGGCGGGGCCCCTGCTGCTCTGCGTAGGACCCTAC  720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    671 CGGGCCGCCTGGGTGGTGGCCGGGCGGGGCCCGGCCGGCGGGGCCCCTGCTGCTCTGCGTAGGACCCTAC  730

Qy    721 AACGCCTCCAACGTGGCCAGCTTCCTGTACCCCAATCTAGGAGGCTCCTGGCGGAAGCTG  780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    731 AACGCCTCCAACGTGGCCAGCTTCCTGTACCCCAATCTAGGAGGCTCCTGGCGGAAGCTG  790

Qy    781 GGGCTCATCACGGGCTGCTTGGAGTGTGGTGCTTAATCCGTTGGTGGTGACCGGTTACTTGGGA  840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    791 GGGCTCATCACGGGCTGCTTGGAGTGTGGTGCTTAATCCGTTGGTGGTGACCGGTTACTTGGGA  850

Qy    841 AGGGGTCCTGACGAGAGCAGTGTGTGCGGCAAGACGCAAGGGGCAAGTCCCAGAAG  900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    851 AGGGGTCCTGACGAGAGCAGTGTGTGCGGCAAGACGCAAGGGGCAAGTCCCAGAAG  910

Qy    901 TAA  903
          |||
Db    911 TAA  913
```

```
RESULT 5
HSU62631
LOCUS       HSU62631        34791 bp    DNA    linear    PRI 15-NOV-2002
DEFINITION  Human B-cell receptor CD22-B isoform and alternatively spliced
            B-cell receptor CD22-A isoform (CD22) gene, complete cds, complete
            sequence.
ACCESSION   U62631
VERSION     U62631.2  GI:24756897
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (sites)
AUTHORS     Wilson,G.L., Fox,C.H., Fauci,A.S. and Kehrl,J.H.
TITLE       cDNA cloning of the B cell membrane protein CD22: a mediator of B-B
            cell interactions
JOURNAL     J. Exp. Med. 173 (1), 137-146 (1991)
MEDLINE     91086838
PUBMED      1985119
REFERENCE   2 (bases 1 to 34791)
AUTHORS     Wilson,G.L., Najfeld,V., Kozlow,E., Menniger,J., Ward,D. and
            Kehrl,J.H.
TITLE       Genomic structure and chromosomal mapping of the human CD22 gene
JOURNAL     J. Immunol. 150 (11), 5013-5024 (1993)
MEDLINE     93267103
PUBMED      8496602
REFERENCE   3 (bases 1 to 34791)
AUTHORS     Lamerdin,J.E., McCready,P., Adamson,A.W., Burkhart-Schultz,K.,
            Garcia,E., Kyle,A., Ramirez,M., Stilwagen,S., Garnes,J.,
            Danganan,L., Bruce,R., Quan,G., Montgomery,M., Ow,D., Kobayashi,A.,
            Olsen,A.O. and Carrano,A.V.
TITLE       Sequence analysis of a 1 Mb region in human 19q13.1
JOURNAL     Unpublished
REFERENCE   4 (bases 1 to 34791)
AUTHORS     Lamerdin,J.E., McCready,P., Adamson,A.W., Burkhart-Schultz,K.,
            Garcia,E., Kyle,A., Ramirez,M., Stilwagen,S., Garnes,J.,
            Danganan,L., Bruce,R., Quan,G., Montgomery,M., Ow,D., Kobayashi,A.,
            Olsen,A.O. and Carrano,A.V.
TITLE       Direct Submission
JOURNAL     Submitted (28-JUN-1996) Human Genome Center, Biology and
            Biotechnology Research Program, Lawrence Livermore National
            Laboratory, 7000 East Ave, Livermore, CA 94550, USA
REFERENCE   5 (bases 1 to 34791)
AUTHORS     Myers,R.M.
TITLE       Direct Submission
JOURNAL     Submitted (07-NOV-2002) DOE Joint Genome Institute, 2800 Mitchell
```

```
REFERENCE   6 (bases 1 to 34791)
AUTHORS     DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE       Direct Submission
JOURNAL     Submitted (14-NOV-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Nov 7, 2002 this sequence version replaced gi:1498644.
            NOTE: This insert is not the entire sequence of the clone (entire
            sequence is 40kb). It is clipped at the overlap with AC002511. The
            number of bases overlapped is 6888.
FEATURES             Location/Qualifiers
     source          1..34791
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /map="19q13.1 betweeen D19S208 and CANPS"
                     /cell_line="UV5HL9-5B"
                     /cell_type="fibroblast"
                     /clone_lib="cosmid f16632; LL19NC02 F2 chromosome
                     19-specific cosmid library"
                     /note="constructed at LLNL from flow-sorted chromosomes
                     from hybrid UV5HL9-5B, which carries chromosome 19 as its
                     only human chromosome"
     repeat_region   complement(128..279)
                     /note="88% similarity"
                     /rpt_family="ALU"
     repeat_region   complement(2035..2277)
                     /note="76% similarity"
                     /rpt_family="ALU"
     repeat_region   complement(2544..2843)
                     /note="85% similarity"
                     /rpt_family="ALU"
     repeat_region   complement(2896..3191)
                     /note="88% similarity"
                     /rpt_family="ALU"
     repeat_region   4547..4848
                     /note="88% similarity"
                     /rpt_family="ALU"
     repeat_region   complement(5422..5707)
                     /note="85% similarity"
                     /rpt_family="ALU"
     repeat_region   6073..6351
                     /note="88% similarity"
                     /rpt_family="ALU"
     gene            6537..21193
                     /gene="CD22"
     mRNA            join(<6537..6570,7043..7420,10532..10837,12251..12517,
                     12664..12927,15377..15634,15839..16102,16198..16461,
                     19325..19421,19547..19622,20098..20216,20647..20731,
                     21062..>21193)
                     /gene="CD22"
                     /product="B-cell receptor CD22-B isoform"
     mRNA            join(<6537..6570,7043..7420,10532..10837,12251..12517,
                     12664..12927,15377..15634,15839..16102,16198..16461,
                     19325..19421,19547..19622,20098..20216,20647..20731,
                     21062..>21193)
                     /gene="CD22"
                     /product="B-cell receptor CD22-A isoform"
     CDS             join(6537..6570,7043..7420,10532..10837,12251..12517,
                     12664..12927,15377..15634,15839..16102,16198..16461,
                     19325..19421,19547..19622,20098..20216,20647..20731,
                     21062..21193)
                     /gene="CD22"
                     /note="alternatively spliced; B-lymphocyte cell adhesion
                     molecule; BL-CAM; human differentiation antigen; B-cell
                     receptor CD22 precursor (alpha and beta)"
                     /codon_start=1
                     /product="B-cell receptor CD22-B isoform"
                     /protein_id="AAB06448.1"
                     /db_xref="GI:1498645"
                     /translation="MHLLGPWLLLLVLEYLAFSDSSKWVFEHPETLYAWEGACVWIPC
                     TYRALDGDLESFLFHNPEYNKNTSKFDGTRLYESTKDGKVPSEQKRVQFLGDKNQNC
                     TLSIHPVHLNDSGQLGLRMESKTEKWMERIHLNVSERPFPHIQLPPEIQESQEVTLT
                     CLLNPSCYGYPIQLQWLLEGVPMRQAAVTSLTIKSVFTRSELKFSPQWSHHGKIVT
                     CQLQDADGKFLSNDTVQLNVKHTPKLEIKVTPSDAIVREGDSVTMCEVSSSNPEYT
                     VSWLKDGTSLKKQNTFTLNLREVTKDQSGKYCCQVSNDVGPGRSEEVFLQVQYAPEPS
                     TVQILHSPAVEGSQVEFLCMSLANPLPTNYTWYHNGKEMQGRTEEKVHIPKILPWHAG
                     TYSCVAENILGTGQRGPGABLDVQYPPKKVTVIQNPMPIREGDTVTLSCNYNSSNPS
                     VTRYEWKPHGAWEEPSLGVLKIQNVGWDNTTIACAACNSWCSWASPVALNVQYAPRDV
                     RVRKIKPLSEIHSGNSVSLQCDFSSSHPKEVQFFWEKNGRLLGKESQLNFDSISPEDA
                     GSYSCWVNNSIGQTASKAWTLEVLYAPRRLRVSMSPGDQVMEGKSATLTCESDANPPV
                     SHYTWFDWNNQSLPYHSQKLRLEPVKVQHSGAYWCQGTNSVGKGRSPLSTLTVYYSPE
                     TIGRRVAVGLGSCLAILLAICGLKLQRRWKRTQSQGLQENSSGQSFFVRNKKVRRA
                     PLSEGPHSLGCYNPMMEDGISYTTLRFPEMNIPRTGDAESSEMQRPPPDCDDTVTYSA
                     LHKRQVGDYENVIPDFPEDEGIHYSELIQFGVGERPQAQENVDYVILKH"
     CDS             join(6537..6570,7043..7420,10532..10837,15377..15634,
                     15839..16102,16198..16461,19325..19421,19547..19622,
                     20098..20216,20647..20731,21062..21193)
                     /gene="CD22"
                     /note="alternatively spliced; B-lymphocyte cell adhesion
                     molecule; BL-CAM; human differentiation antigen; B-cell
                     receptor CD22 precursor (alpha and beta)"
                     /codon_start=1
                     /product="B-cell receptor CD22-A isoform"
                     /protein_id="AAB06449.1"
                     /db_xref="GI:1498646"
                     /translation="MHLLGPWLLLLVLEYLAFSDSSKWVFEHPETLYAWEGACVWIPC
                     TYRALDGDLESFLFHNPEYNKNTSKFDGTRLYESTKDGKVPSEQKRVQFLGDKNQNC
                     TLSIHPVHLNDSGQLGLRMESKTEKWMERIHLNVSERPFPHIQLPPEIQESQEVTLT
                     CLLNPSCYGYPIQLQWLLEGVPMRQAAVTSLTIKSVFTRSELKFSPQWSHHGKIVT
                     CQLQDADGKFLSNDTVQLNVKHTPKLEIKVTPSDAIVREGDSVTMCEVSSSNPEYT
                     VSWLKDGTSLKKQNTFTLNLREVTKDQSGKYCCQVSNDVGPGRSEEVFLQVQYAPEPS
                     TVQILHSPAVEGSQVEFLCMSLANPLPTNYTWYHNGKEMQGRTEEKVHIPKILPWHAG
                     TYSCVAENILGTGQRGPGABLDVQYPPKKVTVIQNPMPIREGDTVTLSCNYNSSNPSVTR
                     YEWKPHGAWEEPSLGVLKIQNVGWDNTTIACAACNSWCSWASPVALNVQYAPRDVRVR
                     KIKPLSEIHSGNSVSLQCDFSSSHPKEVQFFWEKNGRLLGKESQLNFDSISPEDAGSY
                     SCWVNNSIGQTASKAWTLEVLYAPRRLRVSMSPGDQVMEGKSATLTCESDANPPVSHY
                     TWFDWNNQSLPYHSQKLRLEPVKVQHSGAYWCQGTNSVGKGRSPLSTLTVYYSPETIG
                     RRVAVGLGSCLAILLAICGLKLQRRWKRTQSQGLQENSSGQSFFVRNKKVRRAPLS
                     EGPHSLGCYNPMMEDGISYTTLRFPEMNIPRTGDAESSEMQRPPPDCDDTVTYSALHK
                     RQVGDYENVIPDFPEDEGIHYSELIQFGVGERPQAQENVDYVILKH"
     repeat_region   6837..6961
                     /note="83% similarity"
                     /rpt_family="ALU"
     repeat_region   complement(8077..8345)
                     /note="81% similarity"
                     /rpt_family="ALU"
     repeat_region   9293..9591
                     /note="90% similarity"
                     /rpt_family="ALU"
     repeat_region   9789..10357
                     /note="89% similarity"
                     /rpt_family="ALU"
     repeat_region   11454..11750
                     /note="89% similarity"
                     /rpt_family="ALU"
     misc_feature    join(12252..12472,12715..12736,12746..12807)
                     /gene="CD22"
                     /note="similar to the EST sequence deposited under GenBank
                     Accession Number N79417"
     repeat_region   complement(13119..13424)
                     /note="92% similarity"
                     /rpt_family="ALU"
     repeat_region   13457..13745
                     /note="78% similarity"
                     /rpt_family="ALU"
     repeat_region   13787..14358
                     /note="90% similarity"
                     /rpt_family="ALU"
     repeat_region   complement(14802..15075)
                     /note="87% similarity"
                     /rpt_family="ALU"
     conflict        16280
                     /gene="CD22"
                     /note="amino acid 639 changes from 'Y' to 'H'"
                     /citation=[1]
                     /replace="t"
     conflict        16341
                     /gene="CD22"
```

```
                                                        Qy      901 TAA 903
                                                                    |||
                                                        Db    26948 TAA 26950

                                                   RESULT 6
                                                   LOCUS       AR308751                      903 bp    DNA            linear    PAT 12-JUN-2003
                                                   DEFINITION  Sequence 271 from patent US 6555339.
                                                   ACCESSION   AR308751
                                                   VERSION     AR308751.1  GI:31700280
                                                   KEYWORDS    .
                                                   SOURCE      Unknown.
                                                     ORGANISM  Unknown.
                                                               Unclassified.
                                                   REFERENCE   1  (bases 1 to 903)
                                                     AUTHORS   Liaw,C.W., Behan,D.P. and Chalmers,D.T.
                                                     TITLE     Non-endogenous, constitutively activated human protein-coupled
                                                               receptors
                                                     JOURNAL   Patent: US 6555339-A 271 29-APR-2003;
                                                   FEATURES             Location/Qualifiers
                                                     source          1..903
                                                                     /organism="unknown"
                                                                     /mol_type="genomic DNA"
                                                   ORIGIN

                                                   Query Match              99.5%;  Score 898.2;  DB 6;  Length 903;
                                                   Best Local Similarity    99.7%;  Pred. No. 8.3e-132;
                                                   Matches   900;  Conservative    0;  Mismatches     3;  Indels    0;  Gaps    0;

                                                   Qy      1 ATGGACCTGCCCCCGCAGCTCTCCTTCGGCCTCTATGTGGCCGGCTTTGCGCTGGGCTTC 60
                                                              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                   Db      1 ATGGACCTGCCCCCGCAGCTCTCCTTCGGCCTCTATGTGGCCGGCTTTGCGCTGGGCTTC 60

                                                   Qy     61 CCGCTCAACGTCTGGCCATCCGAGGGCGACGGCCCAGCGCCCGGCTCCGTCTCACCCCCT 120
                                                              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                   Db     61 CCGCTCAACGTCTGGCCATCCGAGGGCGACGGCCCAGCGCCCGGCTCCGTCTCACCCCCT 120

                                                   Qy    121 AGCCTGGTCTACGCCCTGAACCTGGGCTGCTCCGACCTGCTGACAGTCTCTCTGCCC 180
                                                              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                   Db    121 AGCCTGGTCTACGCCCTGAACCTGGGCTGCTCCGACCTGCTGACAGTCTCTCTGCCC 180

                                                   Qy    181 CTGAAGGGCGTGGAGGCGCTAGCCTCCGGGGCCTGGCCTTCTGCGGCCTGTGCCCC 240
                                                              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                   Db    181 CTGAAGGGCGTGGAGGCGCTAGCCTCCGGGGCCTGGCCTTCTGCGGCCTGTGCCCC 240

                                                   Qy    241 GTCTTCGCGGTGGCCCACTTCTTCCCACTCTATGCCGGCGGGGCTTCCGGAGGCCG 300
                                                              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                   Db    241 GTCTTCGCGGTGGCCCACTTCTTCCCACTCTATGCCGGCGGGGCTTCCGGAGGCCG 300

                                                   Qy    301 AGTGCAGGGCCGCTACCTGGGAGCAGCCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCG 360
                                                              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                   Db    301 AGTGCAGGGCCGCTACCTGGGAGCAGCCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCG 360

                                                   Qy    361 TGCTATTCCTGGGGGGTGTGCGGCCCATCTGGGTCTG 420
                                                              |||||||||||||||||||||||||||||||||||||
                                                   Db    361 TGCTATTCCTGGGGGGTGTGCGGCCCATCTGGGTCTG 420

                                                   Qy    421 GTCTTTGGGTTGGAGGCTCCAGGAGGCTGGCTGGACCACAGCAACACCCTCGGCATC 480
                                                              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                   Db    421 GTCTTTGGGTTGGAGGCTCCAGGAGGCTGGCTGGACCACAGCAACACCCTCGGCATC 480

                                                   Qy    481 AACACACCGGTCAACGTCAACGGGGCCTCTGGGACCCGGGCCTCTGCGGGC 540
                                                              ||||||||||||||||||||||||||||||||||||||||||||||||||
                                                   Db    481 AACACACCGGTCAACGTCAACGGGGCCTCTGGGACCCGGGCCTCTGCGGGC 540

                                                   Qy    541 CCGGCCCGCTTCAGCCTCTTTTTCTGCCTTGGCCATCACAGCCTTC 600
                                                              ||||||||||||||||||||||||||||||||||||||||||||||
                                                   Db    541 CCGGCCCGCTTCAGCCTCTTTTTCTGCCTTGGCCATCACAGCCTTC 600

                                                   Qy    601 TGCTACGTGGGCTGCCTGGGCGCTGCCGCACTGGGCCGAAGCTG 660
                                                              |||||||||||||||||||||||||||||||||||||
                                                   Db    601
```

```
                     /citation=[1]
                     /replace="c"
           conflict  16451..16454
                     /gene="CD22"
                     /citation=[1]
                     /replace="tact"
           repeat_region   16468..16676
           repeat_region   complement(16698..16953)
                     /note="imperfect CTTT repeat"

           Query Match             100.0%;  Score 903;  DB 9;  Length 34791;
           Best Local Similarity   100.0%;  Pred. No. 6.8e-133;
           Matches   903;  Conservative     0;  Mismatches     0;  Indels    0;  Gaps    0;

           Qy        1 ATGGACCTGCCCCCGCAGCTCTCCTTCGGCCTCTATGTGGCCGGCTTTGCGCTGGGCTTC 60
                        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
           Db    26048 ATGGACCTGCCCCCGCAGCTCTCCTTCGGCCTCTATGTGGCCGGCTTTGCGCTGGGCTTC 26107

           Qy       61 CCGCTCAACGTCTGGCCATCCGAGGGCGACGGCCCAGCGCCCGGCTCCGTCTCACCCCCT 120
                        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
           Db    26108 CCGCTCAACGTCTGGCCATCCGAGGGCGACGGCCCAGCGCCCGGCTCCGTCTCACCCCCT 26167

           Qy      121 AGCCTGGTCTACGCCCTGAACCTGGGCTGCTCCGACCTGCTGACAGTCTCTCTGCCC 180
                        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
           Db    26168 AGCCTGGTCTACGCCCTGAACCTGGGCTGCTCCGACCTGCTGACAGTCTCTCTGCCC 26227

           Qy      181 CTGAAGGGCGTGGAGGCGCTAGCCTCCGGGGCCTGGCCTTCTGCGGCCTGTGCCCC 240
                        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
           Db    26228 CTGAAGGGCGTGGAGGCGCTAGCCTCCGGGGCCTGGCCTTCTGCGGCCTGTGCCCC 26287

           Qy      241 GTCTTCGCGGTGGCCCACTTCTTCCCACTCTATGCCGGCGGGGCTTCCGGAGGCCG 300
                        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
           Db    26288 GTCTTCGCGGTGGCCCACTTCTTCCCACTCTATGCCGGCGGGGCTTCCGGAGGCCG 26347

           Qy      301 AGTGCAGGGCCGCTACCTGGGAGCAGCCTTCCCCTTGGGCTACCAAGCCTTC 360
                        ||||||||||||||||||||||||||||||||||||||||||||||||||||
           Db    26348 AGTGCAGGGCCGCTACCTGGGAGCAGCCTTCCCCTTGGGCTACCAAGCCTTC 26407

           Qy      361 TGCTATTCCTGGGGGGTGTGCGGCCCATCTGGGTCACCTGGGTCTG 420
                        |||||||||||||||||||||||||||||||||||||
           Db    26408 TGCTATTCCTGGGGGGTGTGCGGCCCATCTGGGTCACCTGGGTCTG 26467

           Qy      421 GTCTTTGGGTTGGAGGCTCCAGGAGGCTGGCTGGACCACAGCAACACCCTCGGCATC 480
                        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
           Db    26468 GTCTTTGGGTTGGAGGCTCCAGGAGGCTGGCTGGACCACAGCAACACCCTCGGCATC 26527

           Qy      481 AACACACCGGTCAACGTCAACGGGGCCCGCTGGCCATCACAGCCTTC 540
                        ||||||||||||||||||||||||||||||||||||||||||||||||
           Db    26528 AACACACCGGTCAACGTCAACGGGGCCCGCTGGCCATCACAGCCTTC 26587

           Qy      541 CCGGCCCGCTTCAGCCTCTTCTCTGCTCTTTTTCTGCCTCCCGGGCACTGGGCCTGGGCCGTTCCGGAGGCCGAAGCTG 600
                        ||||||||||||||||||||||||||||||||||||||||||||||||||||
           Db    26588 CCGGCCCGCTTCAGCCTCTTCTCTGCTCTTTTTCTGCCTCCCGGGCACTGGGCCTGGGCCGTTCCGGAGGCCGAAGCTG 26647

           Qy      601 TGCTACGTGGGCTGCCTGGGCGCTGCCGCACTGGGCCGAAGCTG 660
                        |||||||||||||||||||||||||||||||||||||
           Db    26648 TGCTACGTGGGCTGCCTGGGCGCTGCCGCACTGGGCCGAAGCTG 26707

           Qy      661 CGGGCCGGCTGGGGTGGCCGGGGCGGGCCCTGCGCTGCTCGGCTGCCTAGGACCCTAC 720
                        ||||||||||||||||||||||||||||||||||||||||||
           Db    26708 CGGGCCGGCTGGGGTGGCCGGGGCGGGCCCTGCGCTGCTCGGCTGCCTAGGACCCTAC 26767

           Qy      721 AACGCCTCCAACGTCCAGCTTCCTGTACCCCAATCTAGGAGGCTCTGGCGGAAGCTG 780
                        ||||||||||||||||||||||||||||||||||||||||||||||||
           Db    26768 AACGCCTCCAACGTCCAGCTTCCTGTACCCCAATCTAGGAGGCTCTGGCGGAAGCTG 26827

           Qy      781 GGGCTCATCACGGGTGCCTGGAGTGGTGCTTAATCCGCTGGTGACCGGTTACTTGGGA 840
                        |||||||||||||||||||||||||||||||||||||||||||||||||
           Db    26828 GGGCTCATCACGGGTGCCTGGAGTGGTGCTTAATCCGCTGGTGACCGGTTACTTGGGA 26887

           Qy      841 AGGGGTCCTGGCCTGAAGACAGTGTGTGCGGCAAGAACCGCAAGGGGCAAGTCCCAGAAG 900
                        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
           Db    26888 AGGGGTCCTGGCCTGAAGACAGTGTGTGCGGCAAGAACCGCAAGGGGCAAGTCCCAGAAG 26947
```

```
Qy      1 ATGGACCTGCCCCCGCAGCTCTCCTTCGCGCTCTATGTGGCCGCCTTGGCGTTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGGACCTGCCCCCGCAGCTCTCCTTCGCGCTCTATGTGGCGGCGTTGGCGTTC 60

Qy     61 CCGCTCAACGTCCTGGCCATCCGAGGCGCGACGGCGCCACGGCGCCGTCACCCGT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCGCTCAACGTCCTGGCCATCCGAGGGGCGGAGGGGCCACGGCGCCGTCACCCCC 120

Qy    121 AGCCTGGTCTACGCCCTGAACCTGGGTGCTCCGACGCTCTCCTGCGCC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AGCCTGGTCTACGCCCTGAACCTGGGTGCTCCGACCGTTGCTGACGTCTCCCTGCCC 180

Qy    181 CTGAAGGCGGTGGAGGGCGTAGCCTCTGCCTCGCCGCCTCGCTCCGCCCTG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CTGAAGGCGGTGGAGGGCGTGGCCTCGGCCTCGCTCCGCTCCGCCCTCACTGTGCCCT 240

Qy    241 GTCTTCGCGGTGGCCACTCTTCTCCACTCTATGCCGCGGGGGGCTTCCTGGCCGCTG 300
          ||||||||||||||||||||||        |||||||||||||||||||||||||||
Db    241 GTCTTCGGGGTGGCCACTCTTGTCCACTCTATGCCGCGGGGGGCTTCCTGGCCGCCTG 300

Qy    301 AGTGCAGGCCCGCTACCTGGGAGCGGCCTTCCCCTTGGGCCCTGTGTCACCTGGGTCTG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGTGCAGGCCGCTACCTGGGAGCGGCCTTCCCCTTGGGCCCTGTGTCACCTGGGTCTG 360

Qy    361 TGCTATTCCTGGGGGGTGTGCGCGGCCATCTGGGGCCTCCTGTGTGCCCTGTCACCTGGGCATC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGCTATTCCTGGGGGGTGTGCGCGGCCATCTGGGGCCTCCTGTGTGCCCTGTCACCTGGGCATC 420

Qy    421 GTCTTTGGGTTGGAGGCTCCAGGAGGCTGGTGGACCACAGCAACACCTCCCTGGGCAGCTG 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GTCTTTGTGTTGGAGGCTCCGGGAGGCTGGTGGACCACAGCAACACCTCCCTGGGCATC 480

Qy    481 AACACACGGTCAAGGCTCTGGAGGGCCTGGGACCCCGGCCGTCTGCGC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AACACACGGTCAAGGCTCTGGAGGGCCTGGGACCCCGGCCCTCTGCGGC 540

Qy    541 CCGGCCCGCTTCAGCCTCTCTCCTCTTTTTCTGCCCTTGGCCATCAGAGCTTC 600
          ||||||||||||||||||||||||       |||||||||||||||||||||||
Db    541 CCGGCCCGCTTCAGCCTCTCTCCTCTTTTTTTCTGCCCTTGGCCATCAGAGCTTC 600

Qy    601 TGCTACGTGGGCTGCTCCCGGCACTGGCCCGCTCCGGCTGACGGCACAGGGGAAGCTG 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TGCTACGTGGGCTGCTCCCGGCACTGGCCCGCTCCGGCTGACGGCACAGGGGAAGCTG 660

Qy    661 CGGGGCCGCGCCTGGGTGGCCGGGCGGGGGCCTCCTTGCTCGTCTGCGTAGGACCCTAC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 AGGGGCCGCGCCTGGGTAGCCGGGCGGGCCTCCTTGCTCGTCTGCGTAGGACCCTAC 720

Qy    721 AACGCCTCCAACGTGGCCAGCTTCTGAACCCCAATCTGGGAGGCTCCTGGCCGGAAGCTG 780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 AACGCCTCCAACGTGGCCAGCTTCTGAACCCCAATCTGGGAGGCTCCTGGCCGCGAAGCTG 780

Qy    781 GGGCTCATCACGGGTGCCTGGAGTGGTGCTTAATCCGCTGGTGTGACCGGTTACTTGGGA 840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GGGCTCATCACGGGTGCCTGGAGTGGTGCTCAACCCGCTGGTGTGACCGGTTACTTGGGA 840

Qy    841 AGGGGTCCTGGCCTTGAAGACAGTGTGCGGCCAAGGGGGCAAGTCCCAGAAG 900
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 AGGGGTTCCTGGCCTTGAAGACAGTGTGCGGCCAAGACGCAAGGGAGCACGTCCCAGAAG 900

Qy    901 TAA 903
          |||
Db    901 TAA 903
```

RESULT 8
AB095746
LOCUS   AB095746   903 bp   mRNA   linear   ROD 02-SEP-2003
DEFINITION Mesocricetus auratus mRNA for GPR40, complete cds.
ACCESSION   AB095746
VERSION   AB095746.1 GI:34392456
KEYWORDS   .
SOURCE   Mesocricetus auratus (golden hamster)

RESULT 7
AB095743
LOCUS   AB095743   903 bp   DNA   linear   PRI 02-SEP-2003
DEFINITION Macaca fascicularis gene for GPR40, complete cds.
ACCESSION   AB095743
VERSION   AB095743.1 GI:34392450
KEYWORDS   .
SOURCE   Macaca fascicularis (crab-eating macaque)
ORGANISM   Macaca fascicularis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1
  AUTHORS   Itoh,Y., Hosoya,M., Harada,M., Kawamata,Y., Tanaka,H.,
            Kobayashi,M., Fujii,R., Fukusumi,S., Tanaka,Y., Maruyama,M.,
            Uejima,H., Satoh,R., Kizawa,H., Okubo,S., Ogi,K., Komatsu,H.,
            Matsumura,F., Noguchi,Y., Fukatsu,K., Hinuma,S., Fujisawa,Y., and
            Fujino,M.
  TITLE   Expression of GPR40 in pancreatic cells
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 903)
  AUTHORS   Shoji,F., Shoichi,O., Noguchi,Y., Itoh,Y., Kobayashi,M., Fujii,R.
            and Hinuma,S.
  TITLE   Direct Submission
  JOURNAL   Submitted (08-NOV-2002) Shuji Hinuma, Discovery Research
            Laboratories 1, Takeda Chemical Industries, Ltd.; Wadai 10,
            Tsukuba, Ibaraki 300-4293, Japan (E-mail:Hinuma_Shuji@takeda.co.jp,
            Tel:81-298-64-5035, Fax:81-298-64-5000)
FEATURES   Location/Qualifiers
     source   1..903
                     /organism="Macaca fascicularis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9541"
     CDS   1..903
                     /codon_start=1
                     /product="GPR40"
                     /protein_id="BAC82553.1"
                     /db_xref="GI:34392451"
                     /translation="MDLPPQLSFALYVAAFAIGFPLNVLAIRGARAHARRLTPSLVY
                     ALNLGCSDLLLTVSLPLKAVERALASGAWPLASLCPVFGVAHPAPLYAGGFPLAALSA
                     GRYLGAAPPLGVQAFRRPCYSWGVCAAIWALVLCHLGLVFVLEAPGGWLDHSNTSLGI
                     NTPVNGSPVCLEAWDPASAGPARFSLSLLLFFLPLAITAFCYVGCLRALAHSGLTHRR
                     KLRAAWVAGALLTLLLCVGPYNASNVASFLNPNLGGSWRKLGLIITGAWSVVLNPLVT
                     GYIGRGPGLKTVCAARTQGSTSQK"
ORIGIN

Query Match   93.4%; Score 843.8; DB 9; Length 903;
Best Local Similarity 95.9%; Pred. No. 2.9e-123;
Matches 866; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

```
Db   421 GTCCTCGGCTTGGAGGCTCCCGGAGGCTGCTGCCTCCGGTGCTGCTGCTGGAGGCTGGTGCTGCTGAACACCACCAGCAGCTCCCTGGAATC 480
             |||  ||   |||||  |||||| || ||||| ||||||||||||| ||  |||||||||| ||||| |||||||||| ||||||||||
Qy   421 AACACCACCGGTCAACGGCTCCCGGATGGTTCGCTGGTGCTGCTGGAGCCTGGAGATCCCAACTCTGCCGG 540

Db   481 AACACCACCGGTCAACGGCTCCCGGATGGTTCGCTGGTGCTGCTGGAGCCTGGAGATCCCAACTCTGCCGG 540
             |||||||||| ||||||||||||||||||| |||||||||||| |||| ||||||||| ||||||||||
Qy   541 CCGGCCCGCTTCAGCCTCTCTTCTTTCTGCCCTTGGCCATCACAGCCTTC 600

Db   541 CCTGCCGCCGCCAGTTTCTCATCCTCTCTTCTTCGTCGTCCTGGTCATCACCGCCTTC 600
             || |||  ||  ||||| |||||||||| |||||| |||||| || || ||||||||
Qy   601 TGCTACGTGGCTGCCTCCGGCCACTGAGCCGTCCGGCCACAGGCGCAAGAGCTG 660

Db   601 TGCTACGTGGGTGCCTCCGGGCGGGGTCTCGGCCACTGCGGCTGAGCCACAAACGGAAGCTC 660
             |||||||||| |||||| || || || |||||||||| ||| || || |||| ||||||||
Qy   661 CGGGGCGCCCTGGGTGCCGGCGGCCGGACCCTAC 720

Db   661 AGGGCAGCCTGGGCGGGCAGGGGGCCTTTCACCTGCTCTGGGCCGGAAGCTG 720
             || ||  |||||  ||  |||||||| || || || ||| ||||||||| |||
Qy   721 AACGGCTCCAACGTGGCAGCTTCCTGTACCCCAATCTAGGAGGTCCTGGCGGAAGCTG 780

Db   721 AATGCCTCCAATGTGCAGAGTTTCGTAAACCCGGACCTGGGAGGGTCCTGGAGGAGAAGCTG 780
             || ||||| || ||||| || ||| |||||| || ||||| ||||||| || |||| ||||
Qy   781 GGGCTCATCACGGGTGGTGCTCTGGAGTGTGGTGCTTAATCCGGCCGTTACTTGGGA 840

Db   781 GGGCTCATCACAGGGTCTGGAGTGTGGTACTCAACCCGCTGTCACCCGGTTACTTGGGA 840
             |||||||||| |||| ||||| ||||| ||| ||||| |||||| |||||||||||||
Qy   841 AGGGGTTCTGCCTGAAGACAGTGTGCGGCAAGAACGCAAGGGGGCAAGTCCCAGAAG 900

Db   841 GCAAGTCCTGCCTGAAGGACAGCAGTATGTACGACACGGACCAAGGAGGAACAATTCAGAAG 900
             || ||||||||||| || |||| ||| |||||||| || || || || ||||| |||||||
Qy   901 TA 902

Db   901 TA 902
```

```
Qy     1 ATGGACCTGCCCCCGCCCCCAGCTCTCTTCGGCCTCCTATGTGGCCCCTTTGCCCCTTCGGCTTC 60
           |||||||||||  || || ||| ||||| || ||| || |||||| || |||||||| |||||
Db     1 ATGGACCCCGTCTCCCAACTCTTCTTCGCCCTCTATGTGTCTGCCTTCGCGCTGGGCTTC 60

Qy    61 CCGCTCAAGTCCTGCCATCCGAGGCGCGACAGGCGCAGCCCCGGCTCCGTCACCCCT 120
           ||| || ||  ||| ||| || |||||| |||||||||||| || ||||||| ||||
Db    61 CCGCTGAAACCTGTTGGCCATCCGAGGCGCCCGGTCTGCCGTGAAGGCTCGGGCCCCGTCCACCCCC 120

Qy   121 AGCTCGGTTCACGCCCTGAACCTGGGCGTGCTCCGACCTGCTGCACAGTCTCTCTGCCC 180
           |||||  |||| || |||| || |||||| || ||||| ||| ||||| ||| ||||||
Db   121 AACCGGTCTATACACTCCACCTGGCCTGCTCTGACCTGCTCCTGGCCATCACGCTACCC 180

Qy   181 CTGAAGGCGGTGGAGGCGCTAGCCTCCGGGGCCTGGCCTCCGGCCCTGTGCCCC 240
           ||||| |||||||||||||  || ||||||||||| ||| ||  || ||||||||
Db   181 GTGAAGGCCGTGGAGGCCCTGGCTTCTGGGGCCTGGCCGCTGCCCCTGTGCCCT 240

Qy   241 GTCTTCGCGGTGCCCCACTTCTCCCCACTCTATGCGGCGGGGGGGTTCCTGGCCCCTG 300
           |||||| |||| ||| ||||| |||||||| |||| ||||  ||||||  || |||
Db   241 GTCTTTGTCCTGGTCCACTTCGCACTTCTATGCGGGGGGGGGGTTCCTGGCCGCTCTC 300

Qy   301 AGTGCAGGCCGCTACCTGGGAGGAGCAGCCTTCCCGGGCTACCAAGCCTTCCGGAGGCCG 360
           |||  ||||| || |||||| || ||| || ||||| ||||| |||| || |||| |||||
Db   301 AGTGCTGGCCGCTACCTGGGGGCGGCTCCCTTCGGGTACCAGGCCGTTCGCCGGCCGCGC 360

Qy   361 TGCTATTCCTGGGGGGTGTCGCGGCGGCCATCTGGGCCCTCGTGCTCTGTCACTCGGGTCTG 420
           |||||||| |||||| || | ||| ||||||| ||||| ||| ||||| ||||| |||||
Db   361 CGCTACTCCTGGGGCGTGTGTGTGGCTATATGGGCCCTCGTGCTCTGTCACATGGGGCTG 420

Qy   421 GTCTTGGGTTGGGTTGAGGCTCCAGGAGGCTGGCTGAGCCTCGGGTCC 480
           ||| ||||| |||||||||||| |||  ||  || || || || ||||
Db   421 GTCCTCGGCTTGGAGGCTCCCGGAGGCTGCTGCCTCCGGTGCTGCTGCCATC 480
```

Best Local Similarity 79.4%; Pred. No. 1.3e-85;
Matches 716; Conservative 0; Mismatches 186; Indels 0; Gaps 0;

```
Qy    1 ATGGACCTGCCCCCGCAGCTCTCCTCGGCCTCTATGTGGCCGCCCTTGCGCTGGCTTC 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGGACCTGCCCCCACAGCTTCTCCTCGCTCTCTATGTATCGCCTTTGCGCTGGGCTTT 60

Qy   61 CCGCTCAACGTCCTGGCCATCCGGGCCGAGGCGGACGGCGCCCAGCGCCTCCGTCTCACCCCT 120
        ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   61 CCATTGAACTTGTTAGCCATCCGAGGCGCAGTGTCCACGCTAAACTGGCAACTGCAGTCCCC 120

Qy  121 AGCCTGGTCTACGCCCCTGAACCTGGGCTGCTGTCTGACAGTCTTCTTGCCC 180
        ||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db  121 AGCTTGGTCTACACTCCCATCCATCTGGGCTGCCATCCACTCTGCCC 180

Qy  181 CTGAAGGCGGTGAGGCGCTAGCCTCCGGGGCGCTGGCCTTCGTGGCCCGCCGCCG 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 CTGAAGGCTGTGGAGGCCCCTGGCTTCGAGGCCTGGCCCCCTTCTGCCCA 240

Qy  241 GTCTTCGCGGTGGCTGGCCCACTTCTTCCCACTCATGCCGGGGCTTCTTCGGCCCGCCCTG 300
        ||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db  241 GTCTTTGCCTTGGCCCACTTTGCTCCCCCTCTACGCAGGGCGGAGGCTTCCTAGTGCTCTC 300

Qy  301 AGTGCAGGCCGCTACCTGGAGACAGCCTTCCCCTTCGGGCTACCAAGGCCTTCCGGAGGGCCG 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 AGCGCTGGCCGCTACCTGGGGGGCTGCCTGGGGCTACCAAGCCATCCGGAGGGCCC 360

Qy  361 TGCTATTCTGGGGGGTGTGCCGGCCATCTGGGGCCTCGTGTCTGGGTCTG 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 CGCTATTCTGGGGGGTGTGTGTGCTATATGGGCCTGTTCCTCTGCGGGTG 420

Qy  421 GTCTTTGGGTTGGAGGCTCCAGGAGGGCTGCTGGACCTGCGGGACGAACACTTCCCTGGCATC 480
        ||||||||||| | |||||||||||||||||||||||||||||||||||||
Db  421 GCCCTTGGCTTGGAGACTTCCGGAAGCTGGCTGCTGCTGCAACAGTACCAGTTCCCTGGGCATC 480

Qy  481 AACACACCGGTCAACGGCTCTCCGTCTCCGGAGGCCTGGGACCCGGCCTCTGCCGGC 540
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 AACATACCCGTGAATGGCTCACCGGTCTGCCCGGAGAGCCTGGGGATCCCCGGAC 540

Qy  541 CCGGCCCGCTTCAGCCTCTTCCTCTTTTTTCTGCCCATCAGCAGCCCTTC 600
        |||||||||||||||||||||| ||| |||||||||||||||||||||||||
Db  541 CCTGCCCGTCTCAGTTTCTCCATTCTGCTCTTTTCTGCCCTTGGTCATCAGCAGCCCTTC 600

Qy  601 TGCTACGTGGGCTGCTCCGGGCACTGGCCCGCTCCGGGCCTGACGCACAGGCCGAAGCTG 660
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 TGCTAGTGGGGCTGCTCCGGGCCCCTGGTGCGCTCAGGCCTGAGCGCACCAAAACGGAAGCTC 660

Qy  661 CGGGCGCCGCTCGGGTGGCCGGCGGGGCCCTCCTCACGCGCTGCGGAGACCCTAC 720
        ||||||||||||| |||| |||||||||||||||||||||||||||||||||
Db  661 AGGGCAGCTTGGGGTGGCCGGAGGGGCTCTCTGTAGCCTGCTCCTCCACACTCTGCGGGCCCTAT 720

Qy  721 AACGGCCTCCAAACGTGGCCAGCTGGCCAGCTTCCTGTACCCCAATCTAGGAGGCTCCTGGCGGAAGCTG 780
        |||||||||||||||||||| | |||||| ||||||||||||||||||||||
Db  721 AATGCCTCCAATGTGGCTAGTTTCATAAAACCCGGACCTAGGAGGGCTCCTGGGAGGAAGTTG 780

Qy  781 GGGCTCATCACGGGTGCCTGGAGTGTGGTGCCTTAATCCGCTGGGTTACCGGTTACTTGGGA 840
        ||||||||||||||||||||||||| ||||| | |||||||||||||||||||||
Db  781 GGACTCATCACAGAGGGGCCTGGAGTGGTGGTGGTACTCAACCCACTGGTTACTGGTTACTTGGGA 840

Qy  841 AGGGGTCCTGGCCTGAAAGACAGTGTGCGGCGGACAAGTCCAGGGCCAAGTCCCAGAAG 900
        ||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db  841 ACAGGTCCTGGACGGGAACAATATGTGACGGGACTCAAAGAGGAACAATTCAGAAG 900

Qy  901 TA 902
        ||
Db  901 TA 902
```

RESULT 10
AB095745
LOCUS   AB095745   903 bp   mRNA   linear   ROD 02-SEP-2003
DEFINITION   Mus musculus mRNA for GPR40, complete cds.
ACCESSION   AB095745

VERSION   AB095745.1 GI:34392454
KEYWORDS   .
SOURCE   Mus musculus (house mouse)
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Itoh,Y., Hosoya,M., Harada,M., Kawamata,Y., Tanaka,H.,
             Kobayashi,M., Fujii,R., Fukusumi,S., Tanaka,Y., Maruyama,M.,
             Uejima,H., Satoh,R., Kizawa,H., Okubo,S., Ogi,K., Komatsu,H.,
             Matsumura,F., Noguchi,Y., Fukatsu,K., Hinuma,S., Fujisawa,Y. and
             Fujino,M.
  TITLE   Expression of GPR40 in pancreatic cells
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 903)
  AUTHORS   Shoji,F., Shoichi,O., Noguchi,Y., Itoh,Y., Kobayashi,M., Fujii,R.
             and Hinuma,S.
  TITLE   Direct Submission
  JOURNAL   Submitted (08-NOV-2002) Shuji Hinuma, Discovery Research
             Laboratories 1, Takeda Chemical Industries, Ltd.; Wadai 10,
             Tsukuba, Ibaraki 300-4293, Japan (E-mail:Hinuma_Shuji@takeda.co.jp,
             Tel:81-298-64-5035, Fax:81-298-64-5000)
FEATURES   Location/Qualifiers
     source   1..903
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
     CDS   1..903
                     /codon_start=1
                     /product="GPR40"
                     /protein_id="BAC82555.1"
                     /db_xref="GI:34392455"
                     /translation="MDLPPQLSFALYVSAPALGFPLNLLAIRGAVSHAKLRLTPSLVY
                     TLHLGCSDLLAITLPLKAVEALASGAWPLPLPFCPVFALAHFAPLYAGGGPLAALSA
                     GRYLGAAFPFGYQAIRRPRYSWGVCVAIWALVLCHLGLAGLETSGSWLDNSTSSLGI
                     NIPVNGSPVCLRAWDPDSARPARLSFSILLFFLPLVITAFCYVGCLRALVRSGLSHKR
                     KLRAAWVAGGALLTLLLCLGPYNASNVASPINPDLGGSWRKLGLITGAWSVVLNPLVT
                     GYLGTGPGRGTICVTRTQRGTIQK"

ORIGIN

Query Match   66.9%; Score 604.4; DB 10; Length 903;
Best Local Similarity 79.4%; Pred. No. 1.3e-85;
Matches 716; Conservative 0; Mismatches 186; Indels 0; Gaps 0;

```
Qy    1 ATGGACCTGCCCCCGCAGCTCTCCTCGGCCTCTATGTGGCCGCCCTTGCGCTGGCTTC 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGGACCTGCCCCCACAGCTTCTCCTCGCTCTCTATGTATCGCCTTTGCGCTGGGCTTT 60

Qy   61 CCGCTCAACGTCCTGGCCATCCGAGGCGGCGACGGCGCCCAGCGCCTCCGGCTCCACCCCT 120
        ||||||||||||||| |||||||||| |||||||||||||||| |||||||||
Db   61 CCATTGAACTTGTTAGCCATCCGAGGCGCAGTGTCCACGCTAAACTGGCAACTGCAGTCCCC 120

Qy  121 AGCCTTGGTCTACGCCGCTACCTGGGCTGCTGGAGCAGCCTTCCCCTTCGGGCTACCAAGGCCG 180
        ||||||||||||| || |||||||||||||||||||||||||||||||||||||
Db  121 AGCGCTGGCCGCTACCTGGGGGGCTGCCTGGGGCTACCAAGCCATCCGGAGGGCCC 180

Qy  181 CTGGGCCTGCGGGCCCTCCTCACGCGCTGCGGATAAAACCCGGTTCTGTCCTCTGGGTCTG 240
        ||||||||||||||||||| |||||||||||||||||||||||||||||
Db  181 TGCTATTCTGGGGGGTGTGTGTGCTATATGGGCCTGTTCCTCTGCGGGTG 240

Qy  241 GTCTTTGCCTTGGCCCACTTTGCTCCCCCTCTACGCAGCCGGAGCCGAGGCCATC 480
```

```
                                    /mol_type="genomic DNA"
                                    /strain="BALB/c"
                                    /db_xref="taxon:10090"
                    gene            <535..>1437
                                    /gene="Gpr40"
                    mRNA            <535..>1437
                                    /gene="Gpr40"
                                    /product="GPR40"
                    CDS             535..1437
                                    /gene="Gpr40"
                                    /note="GPCR"
                                    /codon_start=1
                                    /product="GPR40"
                                    /protein_id="AAN03478.1"
                                    /db_xref="GI:22597182"
                                    /translation="MDLPPQFSPALYVSAFALGPLNLLAIRGAVSHAKLRITPSLVY
                                    TLHLGCSDLLAITLPLKAVEALASGAWPLPLPFCPVFALAHFAPLYAGGGFLAALSA
                                    GRYLGAAPFGYQAIRRPRYSWGCVAIWALVLCHLGLALGLETSGSWLDNSTSSLGI
                                    NIPVNGSPVCLEAWDPDSARPARLSFSILLFFLPLVITAFCYVGCLRALIVRSGLSHKR
                                    KLRAAWVAGGALLTLLCLGPYNASNVASFINPDLGGSWRKLGLITGAWSVVLNPLVT
                                    GYLGTGPGRGTICVTRTQRGTIQK"
                    ORIGIN
                    Query Match     66.8%;    Score 602.8; DB 10; Length 1819;
                    Best Local Similarity  79.3%;   Pred. No. 1.9e-85;
                    Matches 715; Conservative 0;  Mismatches 187;   Indels 0;  Gaps  0;

Qy         1 ATGGACCTGCCCCCGCAGCTTCCCTTCGGCGCTCTATGTGGCGCCTTGCGTCGGCGTTC   60
             ||||| |||  || ||||| || || ||||||||  || |||||||| ||||||||||
Db       535 ATGGACCTGCCCCCACAGTTCTCCTTCGCTCTATGATCTGCCTTTGCCGCTGGCTTT   594

Qy        61 CCGCTCAACGTCTGCCATCCGAGGGCGACGGCCCACGCCCGGCTCCGTCCACCCCT  120
             ||| ||| ||||| || || || || ||||||| ||||| ||||||||||||||||
Db       595 CCATTGAACTTGTTAGCCATCCGAGCGCGCAGTGTCCCACGCGTAAACTGCGACTCACTCCC  654

Qy       121 AGCTTGGTCTACGCCTGGGCTGCCTCGCCGAACCTGGCTGCTCCGACCTGCTGCTGACAGTCTCTGCCC  180
             ||||||||||||||||| ||||||  |||| || || ||| || ||| ||||||||||||||||||||
Db       655 AGCTTGGTCTCACACTCTCCATCTCCTACTCGGCCATCACTCTGCCC  714

Qy       181 CTGAAGGCGGTGGAGGCGCTAGCCTGGGCTGCCTTGGGGGCGCCGCCGTGTGCCCC  240
             || ||||| || ||||| || || || ||||| ||||||||||||||||||||||
Db       715 CTGAAGGCTGTGGAGGCCCTGGCTTCTGGAGCCTCTATGATCTGCCTTTGCCGCTGGCTGGCTCCGCCA  774

Qy       241 GTCTTCGCCGGTGGCCCACTTCTTCCCCACTCTATGCCGGCGGGGGCTTCCTGGCCGCCCTG  300
             || ||| ||| || || || || ||||| |||| || || || ||||| || ||||||||
Db       775 GTCTTTGCCCTTGGCCCCACTTTGCTCCCCTCCACGCGCAGGCGGCGGAGGGTTCCTAGCTGTGTCTC  834

Qy       301 AGTGCAGGCGCTACCTGGGGGCGTGCGCGGCCCTACCTGGCTGCCGAGGCCG  360
             || |||||| || || |||||| || ||||| || || ||||||||||||||
Db       835 AGCGCTGGCGCTACCTGGGGGGCTGCCTTCGGGTACCAGCCATCCGGAGGCCC  894

Qy       361 TGCTATTCCTGGGGGGTGTGCGCGGCCATCGGGTCCGTGTTCTCGTGTCCTCTGCACCTGGGGTCTG  420
             | || || ||||| ||||| ||||| || || ||| || || || || || || ||||||||||||
Db       895 CGCTATTCCTGGGGTGTGGGTGTGGCTATATGGGCCCTATATGGGGTCTCCACCTGGGGGCTG  954

Qy       421 GTCTTTGGGTTGGAGGCTCCAGGAGGCTGCGTCCAGGAGGCTGCTGACCACCTCCCTGGGCATC  480
             || || |||||||||||| ||||||||||||| ||| ||||||||| || ||||||||||||||
Db       955 GCCCCTGGCTTGGAGACTTCCGGAAGCTGGCTGGTGGCAACAGTACCAGTTCCCTGGGCATC  1014

Qy       481 AACACACCGGTCAACGCCGCTCTCCCGGCACTGGACCCGGCCTGGGACCCGGCCTTGCCGGC  540
             |||| |||| || |||||||||| || ||||||| || ||||| ||||||||||||| ||||
Db      1015 AACATACCCGTGAATGGCTCCCCGTGCTGCCTGGAAGCCTGGGATCCGGACTCTGCCCGC  1074

Qy       541 CCGGGCCGCTTCAGCCTCCTCCTG  600
             ||| ||||||||||||| ||| |||
Db      1075 CCTGCCCGTCCCAGTTTCTCCATTCTCCTTCTCCTTGTTCTCCTTCTCCTTGTCATCACTGCCTTC  1134

Qy       601 TGCTACGTGGGCTGCCTCCGGGCACTG  660
             |||||| || ||||||| ||||  |||
Db      1135 TGCTATGTGGGCTGCCTGCGTGCCCTGGCTCGAGGCCTCACAAACGGAGAAGCTC  1194

Qy       661 CGGGGCCCGCTTGGGTGCCGGCGGGGGGCCCCTCCACGCTGCTGCTTCCACGCTGGGGTAGGACCCTAC  720
             |||| ||||| |  | |||| || ||||| || || ||||||||||| | |||||| | |||||||||
Db       421 GCCCTTGGCTTGGAGACTTCCGGAAGCTGGCTGGTGGCAACACCTCCAGTTCCCTGGGCATC   480

Qy       481 AACACACCGGTCAACGCCGCTCTCCCGGCACTGGACCCGGCCTGGGACCCGGC  540
             ||| |||| |||| || ||||| || |||||| ||| ||||| ||||||||||
Db       481 AACATACCCGTGAATGGCTCCCCGTGCCTGGAAGCCTGGGATCCGGACTCTGCCCGC  540

Qy       541 CCGGGCCGCTTCAGCCTCCTCCTG  600
             ||| |||  |||||||| || |||
Db       541 CCTGCCCGTCCCAGTTTCTCCATTCTCTGCTCTTTGTTCTCTTTGTCATCACTGCCTAT  600

Qy       601 TGCTACGTGGGCTGCCTCCGGGCACTG  660
             |||||| || ||||||| |||| ||||
Db       601 TGCTATGTGGGCTGCCTGCGTGCCCTGGCTCGAGGCCTCACACAGCGGAAGTG  660

Qy       661 CGGGGGCCGCTTGGGTGCCGGCGGGGGGCCCCTCCACGCTGCTGCTTCACGCTGGGGT  720
             |||  | ||| |  | |||| || ||||| ||   |||||||| || |   ||||||
Db       661 AACGCCTCCAACGTGGCCTCGTTTATCAACCCGGACCCTGGTGGTTCCTGGAGGAAGTTG  720

RESULT 11
AF539809
LOCUS       AF539809                1819 bp    DNA     linear   ROD 24-MAR-2003
DEFINITION  Mus musculus GPR40 (Gpr40) gene, complete cds.
ACCESSION   AF539809
VERSION     AF539809.1  GI:22597181
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1819)
  AUTHORS   Andrews,J.L., Briscoe,C.P., Ignar,D.M., Muir,A.I., Sauls,H.R. Jr.
            and Tadayyon,M.
  TITLE     Method of Screening for GPR40 Ligands
  JOURNAL   Patent: PCT WO/02/057783-B 25-JUL-2002;
            Glaxo Group Limited; Glaxo Wellcome House, Berkely Avenue;
            Greenford, Middlesex;
            U.K.;
REFERENCE   2  (bases 1 to 1819)
  AUTHORS   Briscoe,C.P., Tadayyon,M., Andrews,J.L., Benson,W.G.,
            Chambers,J.K., Eilert,M.M., Ellis,C., Elshourbagy,N.A., Goetz,A.S.,
            Minnick,D.T., Murdock,P.R., Sauls,H.R. Jr., Shabon,U.,
            Spinage,L.D., Strum,J.C., Szekeres,P.G., Tan,K.B., Way,J.M.,
            Ignar,D.M., Wilson,S. and Muir,A.I.
  TITLE     The orphan G protein-coupled receptor GPR40 is activated by medium
            and long-chain fatty acids
  JOURNAL   J. Biol. Chem. 278 (13), 11303-11311 (2003)
   PUBMED   12496284
REFERENCE   3  (bases 1 to 1819)
  AUTHORS   Ellis,C. and Elshourbagy,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2002) Metabolic Diseases, GlaxoSmithKline, 5
            Moore Drive, Durham, NC 27709, USA
FEATURES             Location/Qualifiers
     source          1..1819
                     /organism="Mus musculus"
```

```
Db   1195 AGGGCAGCTTGGGTGCCGGAGGCCTCTCCTCACACTCCTGCCTCGGGGCCCTAT 1254
               ||||||||||||||||  ||||   ||||||| || ||||||||||
Qy    721 AACGGCTCCAACGTGGCCAGCTTCCTGTACCCCAATCTAGGAGGCTCCTGCGGAAGCTG 780
               |||||||| ||| |||   |||  |||| ||||||||||||||||||
Db   1255 AATGCCTCCAATGTGCGCTAGTTTCATAAACCGGGACCTAGGAGGCTCCTGGAGGAAGTTG 1314
Qy    781 GGGCTCATCACGGGGTGCCTGGAGTGTGGTGCTTAATCCGCTGGTGACCGGTTACTTGGGA 840
               ||||||||||| |||||||||||||||||  ||    |  ||||| |||||||||||
Db   1315 GGACTCATCACAGGGGCCTGGAGTGTGGTACTCACCCACTGTGGACTTGCTACTTGGGA 1374
Qy    841 AGGGGTCCTGCCTGAAGACAGTGTGTGCGGCAAGAACGCAAGGGGCAGTCCCAGAAG 900
               | ||||||| |||| ||    |  | ||| ||||  |||||||||     |||||
Db   1375 ACAGGTCCTGGACGGGGAACAATATGTGACGAGGACTCAAAGAGGAACAATTCAGAAG 1434
Qy    901 TA 902
               ||
Db   1435 TA 1436
```

RESULT 12
AC087143/c

```
LOCUS       AC087143        209873 bp    DNA     linear   HTG 09-DEC-2000
DEFINITION  Mus musculus clone RP23-22G9, WORKING DRAFT SEQUENCE, 21 unordered
            pieces.
ACCESSION   AC087143
VERSION     AC087143.1  GI:11610869
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 209873)
AUTHORS     DOE Joint Genome Institute.
TITLE       Sequencing of Mouse
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 209873)
AUTHORS     DOE Joint Genome Institute.
TITLE       Direct Submission
JOURNAL     Submitted (09-DEC-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     ------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov

            Project Information
            Center Project Name: 1745473
            Center clone name: RPCI-23_22G9

            Summary Statistics
            Consensus quality: 180765 bases at least Q40
            Consensus quality: 197399 bases at least Q30
            Consensus quality: 200208 bases at least Q20
            Estimated insert size: 186000; agarose-fp estimation
            Estimated insert size: 207873; sum-of-contigs estimation
            Quality coverage: 8.28 in Q20 bases; agarose-fp estimation
            Quality coverage: 7.41 in Q20 bases; sum-of-contigs estimation
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 21 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1853: contig of 1853 bp in length
            *     1854     1953: gap of unknown length
            *     1954     2957: contig of 1004 bp in length
            *     2958     3057: gap of unknown length
            *     3058     4549: contig of 1492 bp in length
            *     4550     4649: gap of unknown length
            *     4650     6366: contig of 1717 bp in length
            *     6367     6466: gap of unknown length
            *     6467     8409: contig of 1943 bp in length
            *     8410     8509: gap of unknown length
            *     8510    10529: contig of 2020 bp in length
            *    10530    10629: gap of unknown length
            *    10630    13145: contig of 2516 bp in length
            *    13146    13245: gap of unknown length
            *    13246    16642: contig of 3397 bp in length
            *    16643    16742: gap of unknown length
            *    16743    20544: contig of 3802 bp in length
            *    20545    20644: gap of unknown length
            *    20645    26371: contig of 5727 bp in length
            *    26372    26471: gap of unknown length
            *    26472    30287: contig of 3816 bp in length
            *    30288    30387: gap of unknown length
            *    30388    36287: contig of 5900 bp in length
            *    36288    36387: gap of unknown length
            *    36388    44527: contig of 8140 bp in length
            *    44528    44627: gap of unknown length
            *    44628    55698: contig of 11071 bp in length
            *    55699    55798: gap of unknown length
            *    55799    66947: contig of 11149 bp in length
            *    66948    67047: gap of unknown length
            *    67048    82580: contig of 15533 bp in length
            *    82581    82680: gap of unknown length
            *    82681   100979: contig of 18299 bp in length
            *   100980   101079: gap of unknown length
            *   101080   124586: contig of 23507 bp in length
            *   124587   124686: gap of unknown length
            *   124687   151243: contig of 26557 bp in length
            *   151244   151343: gap of unknown length
            *   151344   181555: contig of 30212 bp in length
            *   181556   181655: gap of unknown length
            *   181656   209873: contig of 28218 bp in length.
FEATURES             Location/Qualifiers
     source          1..209873
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-22G9"
                     /clone_lib="RPCI mouse BAC library 23"
ORIGIN

Query Match     66.8%;   Score 602.8;   DB 2;   Length 209873;
Best Local Similarity  79.3%;   Pred. No. 7.2e-86;
Matches  715;   Conservative   0;   Mismatches 187;   Indels   0;   Gaps   0;

Qy      1 ATGGACCTGCCCCGCAGCTCTCCTTCGGCCTCATGTGGCCGCCTTGCGTCGGGGCTTC   60
          ||||||||||||| |||||||||| |||||| ||||||| | |||||||||||| ||||
Db 107167 ATGGACCTGCCCCCACAGCTCTCCTTCGTCGTCTCTATGTATCCGCTTGCGCGTGGGGCTTT 107108

Qy     61 CCGCTCAAGCTCCTGGCCATCCGAGCGGCGACGGCGCCCAGCCGTCCGTCTCACCCCT   120
          ||   | ||| ||| |||||||||||||  |||| ||||| ||| ||  ||||||||
Db 107107 CCATTGAACTTGTTAGCCATCCGAGCGGCCAGTGTCCCACGCTAAACTGCGACTCACTCCCC 107048

Qy    121 AGCCTGGTCTACGCCCGGAACCTGGCGTCGGGCGCTCGGCGCTGTCGCTGACAGTCTCTGCCC   180
          |||  ||||||||  | |||| |||||| || |  |||| | |||||| |  ||||||||
Db 107047 AGCTTGGTCTACACTCTCCATCTGGCGCTCGATCCTCGATCTCGTCGCCATCACTCTGCCC 106988

Qy    181 CTGAAGGCGGTGGAGGCGCTAGCCTCCGGGGCCTCCGGGCCTCGTCCGGGCCTCGTGTGCCC   240
          |||||||| |||||||||| || | |||  |||  | ||||||||||| ||||||| |||
Db 106987 CTGAAGGCTGTGGAGGCGCTGGAGCCTCATGCCGCGGCCTGCGGGGGCCTCGCGCGCCGCCGCCA 106928

Qy    241 GTCTTCGCGGTGCCTTGCCCACTCTTCTCCCACTCACTCCCCTCGGGCCTCCGGGCCGCCTG   300
          |||||  || ||||| ||||||| |  ||||  ||  || | |  |||  || ||| |||
Db 106927 GTCTTTGCCTTGCCCTTGCCCACTTTGCTCCCCTTCGCAGAGGCGAGGCGCTTCCTAGCTGCTCTC 106868

Qy    301 AGTGCAGGCCCTACCTGGGAGCAGCAGCTTCCCGTTGGGCTACCAAGCCTTCCGGAGGCCG   360
          || || ||| ||||||||| || ||  |  | |||||||||||||| ||| |||||| ||
Db 106867 AGCGCTGGCGCCTACCTGGGGGCTGCTCGGGTACCAAGCCATCCGGAGAGCCC 106808

Qy    361 TGCTATTCCTGGGGGTGTGCGCGGCCATCTGGGCCCTCGCCTGTGTCACTGGGTCTG   420
          ||||||||||||||||||| | ||| ||| ||||||| | ||| |||| |||| |||
Db 106807 CGCTATTCCTGGGGGTGTGTGTGGCGTATATGGGCCCCTTGTGTCTGCCACCTGGGGGCTG 106748
```

```
Qy    421 GTCTTTGGGTTGGAGGCTGGCTGACCACAGCAACACCTCCCTGGGCATC            480
          |||||| ||||||||| || ||||  |||||||||| |  |  ||||||||||||
Db 106747 GCCCTTGGCTTGGAGACTTCCGGAAGCTGGCTGGACAACAGTACCAGTTCCCTGGGCATC 106688

Qy    481 AACACACCGGTCAACGGCGCTCCCGTCTGCCTGGACCCGCCTCTGCCCGC            540
          |||| ||| || || ||| |||||| ||||||||| ||| |  | |||||||||||
Db 106687 AACATACCCGTGAATGGCTCTCCCCGTCTGCCTGAAGCTCGGGATCCCGACTCTGCCCGC 106628

Qy    541 CCGGCCGCTTCAGCCTCTCTCTCTCCTGCTCTTTTTCTGCCCATCACAGCCTTC       600
          || |||||||||| |||| | |||| |||||| ||| ||||||| ||||| |||||
Db 106627 CCTGCCCGTCTCAGTTTCTCCATTCTGCTCTTCTTCTTTCTGCCCTGGTCATCACTGCCTTC 106568

Qy    601 TGCTACGTGGGCTGCCTCCGGGCACTGACCCGTCCGGCCCGCTGACGCAGGCCGGAAGCTG 660
          ||||| ||||||||||||| |||| |||  ||| | ||||||| ||||| |||| |||||||
Db 106567 TGCTATGTGGGCTGCCTCCCGGGCCCTGGTGCGCTCAGGCGTGAGCCACAAACGGAAGCTC 106508

Qy    661 CGGGCCGCCTGGGTGGGCCGGCGGGGCCCCTCCTCCACGCGTCTGCTCTGGGTAGGACCCTAC 720
          |||||   ||||| || | || ||| | | ||| |||| | | ||||||||     |||||||
Db 106507 AGGGCAGTTTGGGTGGCCGGAGGGCGCTCTCTCCACACTCCTGCTCTGCTCTGGGGCCCTAT 106448

Qy    721 AACGGCCTCCAACGTGGCCAGCTTCCTGTACCCCAATCTAGGAGGCTCCTGGCGGAAGCTG 780
          || | |||||| |||||  || ||| | | ||| || |||||||||||||| |||||| |
Db 106447 AATGCCCTCCAATGTGGCTAGTTTCATAAACCCGGAACCTAGGAGGCTCCTGGAGGAAGTTA 106388

Qy    781 GGGCTCATCACGGGTGCCTGGAGTGTGTGGTGCTTAATCCGCTGGTTACTTGGGA       840
          || |||||||| || ||||||||||||  | || | ||| |||| ||| |||||||
Db 106387 GGACTCATCACAGGGCCCTGGAGTGTGGTACTCAACCCACTGGTCACTGCTACTTGGGA 106328

Qy    841 AGGGGTCCTGGCCTGAAGACAGTGTGTGCGGCAAGAACCAAGGGGCAAGTCCCAGAAG    900
           | |||||||| | |  |  || ||| |||  || ||| ||| ||| | |||||||||
Db 106327 ACAGGTCCTGGACGGGGAACAATATGTGACGAGGACTCAAAGAGAGAACAATTCAGAAG 106268

Qy    901 TA 902
          ||
Db 106267 TA 106266
```

RESULT 13
AC079472

LOCUS   AC079472   154673 bp   DNA   linear   HTG 02-SEP-2000
DEFINITION Mus musculus clone RP23-123D23, WORKING DRAFT SEQUENCE, 34
            unordered pieces.
ACCESSION   AC079472
VERSION   AC079472.1 GI:9964837
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 154673)
  AUTHORS   DOE Joint Genome Institute.
  TITLE   Sequencing of Mouse
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 154673)
  AUTHORS   DOE Joint Genome Institute.
  TITLE   Direct Submission
  JOURNAL   Submitted (02-SEP-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT   -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov

            Project Information
            Center Project Name: 1784199
            Center clone Name: RPCI-23_123D23

            Summary Statistics
            Consensus quality: 129250 bases at least Q40
            Consensus quality: 141366 bases at least Q30
            Consensus quality: 144402 bases at least Q20
            Estimated insert size: 221980; agarose-fp estimation
            Estimated insert size: 151373; sum-of-contigs estimation
            Quality coverage: 9.04 in Q20 bases; agarose-fp estimation
            Quality coverage: 13.25 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 34 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accesion number will
            * be preserved.
            *   1   1048: contig of 1048 bp in length
            *   1049   1148: gap of unknown length
            *   1149   2587: contig of 1439 bp in length
            *   2588   2687: gap of unknown length
            *   2688   3845: contig of 1158 bp in length
            *   3846   3945: gap of unknown length
            *   3946   5134: contig of 1189 bp in length
            *   5135   5234: gap of unknown length
            *   5235   6339: contig of 1105 bp in length
            *   6340   6439: gap of unknown length
            *   6440   7755: contig of 1316 bp in length
            *   7756   7855: gap of unknown length
            *   7856   9410: contig of 1555 bp in length
            *   9411   9510: gap of unknown length
            *   9511   11291: contig of 1781 bp in length
            * 11292   11391: gap of unknown length
            * 11392   13640: contig of 2249 bp in length
            * 13641   13740: gap of unknown length
            * 13741   15136: contig of 1396 bp in length
            * 15137   15236: gap of unknown length
            * 15237   16891: contig of 1655 bp in length
            * 16892   16991: gap of unknown length
            * 16992   18153: contig of 1162 bp in length
            * 18154   18253: gap of unknown length
            * 18254   20654: contig of 2401 bp in length
            * 20655   20754: gap of unknown length
            * 20755   23725: contig of 2971 bp in length
            * 23726   23825: gap of unknown length
            * 23826   26327: contig of 2502 bp in length
            * 26328   26427: gap of unknown length
            * 26428   27978: contig of 1551 bp in length
            * 27979   28078: gap of unknown length
            * 28079   32606: contig of 4528 bp in length
            * 32607   32706: gap of unknown length
            * 32707   36800: contig of 4094 bp in length
            * 36801   36900: gap of unknown length
            * 36901   40299: contig of 3399 bp in length
            * 40300   40399: gap of unknown length
            * 40400   45089: contig of 4690 bp in length
            * 45090   45189: gap of unknown length
            * 45190   49948: contig of 4759 bp in length
            * 49949   50048: gap of unknown length
            * 50049   57292: contig of 7244 bp in length
            * 57293   57392: gap of unknown length
            * 57393   61258: contig of 3866 bp in length
            * 61259   61358: gap of unknown length
            * 61359   66936: contig of 5578 bp in length
            * 66937   67036: gap of unknown length
            * 67037   73385: contig of 6349 bp in length
            * 73386   73485: gap of unknown length
            * 73486   83300: contig of 9815 bp in length
            * 83301   83400: gap of unknown length
            * 83401   90260: contig of 6860 bp in length
            * 90261   90360: gap of unknown length
            * 90361   95971: contig of 5611 bp in length
            * 95972   96071: gap of unknown length
            * 96072   102971: contig of 6900 bp in length
            * 102972   103071: gap of unknown length
            * 103072   115455: contig of 12384 bp in length
            * 115456   115555: gap of unknown length
            * 115556   123966: contig of 8411 bp in length
            * 123967   124066: gap of unknown length

* 124067 132993: contig of 8927 bp in length
* 132994 133093: gap of unknown length
* 133094 142772: contig of 9679 bp in length
* 142773 142872: gap of unknown length
* 142873 154673: contig of 11801 bp in length.

FEATURES Location/Qualifiers
source 1..154673
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-123D23"
/clone_lib="RPCI mouse BAC library 23"

ORIGIN

Query Match 66.2%; Score 598; DB 2; Length 154673;
Best Local Similarity 79.3%; Pred. No. 4.4e-85;
Matches 709; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

Qy   1 ATGGACCTGCCCCCGCAGCTCTCCTTCGGCCTCTATGTGGCCGCCTTTGCGCTGGGCTTC 60
           ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db 114152 ATGGACCTGCCCCCACCAGCTCTCCTTCGCTCCTTCCTATGTATCTGCCTTTGCGCTGGGCTTT 114211

Qy   61 CCGCTCAACGTCCTGGCCATCGGAGGCGGCGACGGCCCAGCCCGGTCCGTCTCACCCCT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 114212 CCATTGAACTTGTTAGCCATCCGAGGCGCAGTGTCCCACCGTAAACTGCGACTCACTCCC 114271

Qy   121 AGCCTGGTCTACGGCCCTGAACCTGGGCTGCTGTCTCCGACCTGCTGCACAGTCTCTCGCCC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 114272 AGCTTGGTCTACACTCTCCATCTGGCTGCTCTGATCCTCCTACTGGCCATCACTCTGCCC 114331

Qy   181 CTGAAGGCGGTGAGGCGGCTAGCCTCCGGGGCCTTCGCGGCCTTCCGGCCTCGCTGTCGCCCC 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 114332 CTGAAGGCGTGGAGGCCCTGGCTTCTGGAGGCTTCCGGGACTACCAGGCCCTCCCCTTCTGCCCA 114391

Qy   241 GTCTTCGCGGTGGCCCACTTCTTCCCACTCTATGCCGGGGGGCTTCCTGGCGCCGCCCTG 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 114392 GTCTTTGCCTTGGCCCACTTGCTCCCCTTGCTCCGGCAGGAGGGCTTCCTAGCTGCTGCTCTC 114451

Qy   301 AGTGCAGGCCGCTACCTGGGAGCAGCCTTCCCCTGGCTACCAAGCCTTCCGGAGGCCG 360
           |||||||||||||||||||||||||||||||||||||||||||||
Db 114452 AGCCGCTGGCCGCTACCTGGGGGCGCTGGGTGCCTTCCCCGTGGGGGACCATCCGGAGGCCC 114511

Qy   361 TGCTATTCCTGGGGGGTGTGGGGGCCATCGGCGGCCATCTGGGCCTCCTGTCACCTGGGTCTG 420
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 114512 CGCTATTCCTGGGGTGTGTGGGGCTGTGGCTATATGGGGCCCTTGTCCCCGCCACCTGGGGCTG 114571

Qy   421 GTCTTTGGGTTGGAGGCTCCAGGAGGCTGGCTGACCAGCAACCAACACCTCCCTGGGCATC 480
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db 114572 GCCCTTGGCTTGGAGACTTCCGGAAGCTGCGTGGGACAGTACCCAGTTCCCTGGGCATC 114631

Qy   481 AACACCACCGGTCAACGGCTCTCCGGTCTGCCTGAGCCCGGAGCACCCGGGCTCTGCCCGGC 540
           |||||||||||||||||||||||||||||||||||||||||||||
Db 114632 AACATACCCGTGAATGGCTTCCCCGTCTGCTGGAAGCCTCCGGGATCCCGGACTCTGCCCGC 114691

Qy   541 CCGGCCCCGCTTCAGCCTTCCCGTTCTCTTTTTCTGCCTCTTCTTTCTTTCTGGTCATCAGCGCTTC 600
           ||||||||||||||||||||||||||||||||||||||||||||
Db 114692 CCTGCCCGTCCGTTCCCAGTTCTCCATCTCGCCATCTCGCCTCTTCTTCTTGGTCATCATCAGCGCCTTC 114751

Qy   601 TGTACGTGGGCTGCCTCCGGACTCACTGGCCCGGCCGCTGACGGCAGGCCGGAAGCTG 660
           ||||||||||||||||||||||||||||||||||||||||||||
Db 114752 TGCTATGTGGGCTGCCTCCGGCCCGGGGCCGGCGCCTCGAGCGCTCGCTCCACAAACGGAAGCTC 114811

Qy   661 CGGGCCGGCTTGGGTGGCCGGGCCGGGCGGGGAGGCGCCTCTGCTCGCTGCCTAGGACCCTAC 720
           |||||||||||||||||||||||||||||||||||||||||
Db 114812 AGGGCAGCTTGGGTGGCCGGGCCGGGCTCTCCTGCTCGCCGCTCGCTGCCTCGGGCCCTAT 114871

Qy   721 AACGCCTCCAACGTGGCCAGCTTCCTGTACCCCGAATCCGACACCCGGACCCGGGAGGCTG 780
           ||| |||||||||||||||||||||||||||||||||||||
Db 114872 AATGCCTCCAATGTGGCTAGTTTCATAAACCCGGACCTAGGAGGGCTCGGGGAGGAAGTTA 114931

Qy   781 GGGTTACGCGGGTGCCTGGAGTGTGGTTGGTACTCAACCCACCTGGTACTGCTACTTGGGA 840
           ||||||||||||||| |||||||||||||||||||||||||||||||
Db 114932 GGGACTCATCACGGGTCCTGGAGTGTGGGTAGTGGGTACTCAACCCACTGGTACTGCTACTTGGGA 114991

Qy   841 AGGGGTTCCTGGCCTGAAGACACAGTGTGTGCGGCAAGAACGCAAGGGGCAAGTCC 894
           ||||||||||||||||||||||||||||||||||||||||||||
Db 114992 ACAGGTCCTGGACGCGGGAACAATATGTGTGACGAGGACTCAAAGAGAGAACATTC 115045

RESULT 14
AB095744
LOCUS   AB095744   903 bp   mRNA   linear   ROD 02-SEP-2003
DEFINITION Rattus norvegicus mRNA for GPR40, complete cds.
ACCESSION   AB095744
VERSION   AB095744.1 GI:3492452
KEYWORDS   .
SOURCE   Rattus norvegicus (Norway rat)
  ORGANISM Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.

REFERENCE   1
  AUTHORS   Itoh,Y., Hosoya,M., Harada,M., Kawamata,Y., Tanaka,H.,
            Kobayashi,M., Fujii,R., Fukusumi,S., Tanaka,Y., Maruyama,M.,
            Uejima,H., Satoh,R., Kizawa,H., Okubo,S., Ogi,K., Komatsu,H.,
            Matsumura,F., Noguchi,Y., Fukatsu,K., Hinuma,S., Fujisawa,Y. and
            Fujino,M.
  TITLE   Expression of GPR40 in pancreatic cells
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 903)
  AUTHORS   Shoji,F., Shoichi,O., Noguchi,Y., Itoh,Y., Kobayashi,M., Fujii,R.
            and Hinuma,S.
  TITLE   Direct Submission
  JOURNAL   Submitted (08-NOV-2002) Shuji Hinuma, Discovery Research
            Laboratories 1, Takeda Chemical Industries, Ltd.; Wadai 10,
            Tsukuba, Ibaraki 300-4293, Japan (E-mail:Hinuma_Shuji@takeda.co.jp,
            Tel:81-298-64-5035, Fax:81-298-64-5000)
FEATURES   Location/Qualifiers
     source   1..903
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
     CDS   1..903
                     /codon_start=1
                     /product="GPR40"
                     /protein_id="BAC82554.1"
                     /db_xref="GI:3492453"
                     /translation="MDLPPQLSFALYVSAFALGPPLNLLAIRGAVSHAKLRLTPSLVY
                     TLHLACSDLLAITLPLKAVEALASGVWPLPLPFCPVFALAHFAPLYAGGGFLAALSA
                     GRYLGAAPFGYAIRRPCYSWGVCVAIWALVLCHLGLAGLEAPRGWVDNTTSSLGI
                     NIPVNGSPVCLEAWDPDSARPARLSFSILLFFPLVITAFCYVGCLRALVHSGLSHKR
                     KLRAAWVAGALLTLLLCLGPYNASNVASPINPDLEGSWRKLGLITGAWSVVLNPLVT
                     GYLGTGPGQGTICVTRTPRGTIQK"

ORIGIN

Query Match 65.3%; Score 590; DB 10; Length 903;
Best Local Similarity 78.4%; Pred. No. 2.3e-83;
Matches 707; Conservative 0; Mismatches 195; Indels 0; Gaps 0;

Qy   1 ATGGACCTGCCCCCGCAGCTCTCCTTCGGCCTCTATGTGGCCGCCTTTGCGCTGGGCTTC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGGACCTGCCCCCACAGCTCTCCTTCGCTCTCTATGTATCAGCGCCTTTGCACTAGGCTTT 60

Qy   61 CCGCTCAACGTCCTGGCCATCGGAGGCGGCGACGGCCCAGCCCGGTCCGTCTCACCCCT 120
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CCATTGAACTTGTTAGCCATCCGAGGCGCAGTGTCCCACCGGAAACTGCGACTCACCCCC 120

Qy   121 AGCCTGGTCTACGGCCCTGAACCTGGGCTGCTGTCTCCGACCTGCTGACAGTCTCTGCCC 180
           |||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AGCTTGGTCTACACTCTCCATTGGCCTCGCTGTCTCCCTCGCTCCTACTGGCCATCACCCTGCCC 180

Qy   181 CTGAAGGCGGTGGAGGCCCTAGCCTCCGGGGCCTTCCGGCCTCGCCGCCTGTCGCCC 240
           ||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CTGAAGGCTGTGGAGGCCCTTGCTTCTGGGGTCTGGCCCCTTCCGCTCCCACTCCCTTCTGCCCA 240

Qy   241 GTCTTCGCGGTGGCCCCACTTCTTCCCACTCTATGCCGGCGGGGGCCTTCCTGGCGGGGCCCTG 300

JOURNAL   Unpublished
REFERENCE 3 (bases 1 to 907)
AUTHORS   Ellis,C.E. and Elshourbagy,N.
TITLE   Direct Submission
JOURNAL   Submitted (01-AUG-2002) Metabolic Diseases, GlaxoSmithKline, 5
           Moore Drive, Durham, NC 27709, USA
FEATURES   Location/Qualifiers
     source   1..907
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
     gene   1..907
                     /gene="Gpr40"
     CDS   5..907
                     /gene="Gpr40"
                     /note="GPCR"
                     /codon_start=1
                     /product="GPR40"
                     /protein_id="AAN03479.1"
                     /db_xref="GI:22597184"
                     /translation="MDLPPQLSFALYVSAFALGPPLNLLAIRGAVSHAKLRLTPSLVY
                     TLHLACSDLLAITLPLKAVEALASGVWPLPLPFCPVFALAHFAPLYAGGGFLAALSA
                     GRYLGAAPPFGYQAIRRPCYSWGVCVAIWALVLCHLGLALGLEAPRGWVDNTTSSLGI
                     NIPVNGSPVCLEAWDPDSARPARLSFSILLFFLPVITAFCYVGCLRALVHSGLSHKR
                     KLRAAWVAGGALLTLLLCLGPYNASNVASFINPDLEGSWRKLGLITGAWSVVLNPLVT
                     GYLGTGPGQGTICVTRTPRGTIQK"

ORIGIN

Query Match   65.3%;   Score 590; DB 10; Length 907;
Best Local Similarity 78.4%; Pred. No. 2.3e-83;
Matches 707; Conservative   0; Mismatches 195; Indels   0; Gaps   0;

Qy   1 ATGGACCTGCCCCCAGCTCTCCTTGCGCTCTATGGCCGGGCCGTGTCGCACGCCAAGCTTC 60
           ||||||||||| || |||||||||||||||| | ||||| | |||||| ||||||||
Db   5 ATGGACCTTGCCTTGCCCCCACCAGCTCTCGCTCTCATGTCTATCAGCGTTTTGCACTAGGCTTT 64

Qy   61 CCGCTCAACGTCCTGCCATCCGAGGCGGCGACGGCCCACGCCGTCGTGCACCCCT 120
           ||||| |||||| | |||||||||||||||||| |||||| || ||||||||
Db   65 CCATTGAACTTGTTAGCCATCCGAGGTGCAGTGTCCCACGCGAAACTGCCACCCCC 124

Qy   121 AGCTGGTCTACGCCCTGGCTCGGCCTCGACCTGGCCTCCGAACCTGGCTGCTGACAGTCTCTGCCC 180
           |||||||||| ||||||||||||||||||||||||| ||||||||||| |||||||||||||||
Db   125 AGCTTGGTCTCACACTCTCCATTGGCCTCGCTCGAGCCTCCTGACACCTCCTCCTCTGCCC 184

Qy   181 CTGAAGGCGGTGGAGGCGTAGCCTGGCCCGGCCCTGCCCCCACTCCCTGTGCCC 240
           ||||||||||||| || ||||||||||||||||| || ||| |||||||| ||||
Db   185 CTGAAGGCGGTGGAGGCGTAGCCTGGCCCGGCCCTTCTGGGGGGTCTGGGGGGCTTCCCCCA 244

Qy   241 GTCTTCGCGTGGCGGTGGCCCACTTCTTCCCACTCTATGGCCGGCCTGTGCCGGGGGGGCTTCCTGGCCGGGCTG 300
           ||||||||||||||| |||||||||||||||||||| || ||||||| ||||||
Db   245 GTCTTTGCCTTGCCCCCACCAGCTCTCGCTCTCATGTCTATCAGCGTGGAGGGCTCCGCTGCTC 304

Qy   301 AGTGCAGGGCGCTACCTGGGAGCGGACGGCGGACGGCCCACGCCGTCGTGCACCCCT 360
           |||||||||||||||| ||||| || |||||||||||||| ||||||||||||
Db   305 AGTGCTGGGCGCTACCTGGGAGCTGGCCCATCCGAGGCCG 364

Qy   361 TGCTATTCCTGGGGGTGTGCGCGGCCATCTGGGCCCTGGTCACCTGGGTCTG 420
           ||||||||||||| ||||||||||||||||||||||| || |||||||||
Db   365 TGCTATTCCTGGGGTGTGCGCGGCCATCGGGCTGGCTATATGGCCCACCTGGGACTG 424

Qy   421 GTCTTTGGGTTGGAGGCTCAGGAGGCTGCTGGCCGACCACAGCAACACCTCCTGGGCATC 480
           |||||| ||| ||||| ||||||||||||| |||||||||| ||| ||||||||
Db   425 GCTCTTGGCTTGGAGGCTCCCAGAGCGTGGGTGGATAACACCACCAGTTCCTGGGCATC 484

Qy   481 AACACACCGGTCAACGGCTTCGAGGCCTTGGACCCGGCCTGGACCTGGTCTG 540
           |||| ||||||||||||||||||||||||||||||||||| || ||||||||
Db   485 AACATACCCGTGAATGGCTTCCCCCGGTCTCCCGTGGGATCTGGAAGCGTGGAGCGTGGGATCTGGACTGCCGC 544

Qy   541 CCGGGCCCGCTTCAGCCTGCTCTCTTTTTCTGCGTCTTTTTCTGGCCATCACAGCGTTC 600
           ||||||||||| ||||||||||||| | ||||||||| |||||||||| | ||||||||
Db   545 CCTGCCCGACTCAGTTCTCGATTCGCTCTTCCTTGTTTATCACTGTTTC 604

RESULT 15
AF539810
LOCUS   AF539810   907 bp   mRNA   linear   ROD 24-MAR-2003
DEFINITION Rattus norvegicus GPR40 (Gpr40) mRNA, complete cds.
ACCESSION   AF539810
VERSION   AF539810.1 GI:22597183
KEYWORDS   .
SOURCE   Rattus norvegicus (Norway rat)
  ORGANISM Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 907)
  AUTHORS   Briscoe,C.P., Tadayyon,M., Andrews,J.L., Benson,W.G.,
            Chambers,J.K., Eilert,M.M., Ellis,C., Elshourbagy,N.A., Goetz,A.S.,
            Minnick,D.T., Murdock,P.R., Sauls,H.R. Jr., Shabon,U.,
            Spinage,L.D., Strum,J.C., Szekeres,P.G., Tan,K.B., Way,J.M.,
            Ignar,D.M., Wilson,S. and Muir,A.I.
  TITLE   The orphan G protein-coupled receptor GPR40 is activated by medium
            and long-chain fatty acids
  JOURNAL   J. Biol. Chem. 278 (13), 11303-11311 (2003)
   PUBMED   12496284
REFERENCE   2 (bases 1 to 907)
  AUTHORS   Ellis,C.E. and Elshourbagy,N.
  TITLE   Rat GPR40 cDNA

(Left-side alignment fragments:)

Db   241 GTCTTTGCCTTGCCCCACTTGCGCCCCTTGCCCCCACCAGCAGCTGGAGGCTTCCTGGCTGCTCTC 300
           |||||||| || |||||||||||||||||||| | ||||||||||||||
Qy   301 AGTGCAGGGCGCTACCTGGGAGCGGACGGCGGACGGCCCACGCCGTCGTG 360
           |||||| ||||||||||||||||||||| ||||||||| ||||||||||||
Db   301 AGTGCTGGGCGCTACCTGGGAGCTGGCCCATCCGAGGCCC 360

Qy   361 TGCTATTCCTGGGGGTGTGCGCGGCCATCTGGGCCCTGGTCACCTGGGTCTG 420
           |||||||||||| ||||||||||||||||| ||| |||||||||||
Db   361 TGCTATTCCTGGGGTGTGCGCGGCCATCTGGGCTGGCTATATGGGGCCATC 420

Qy   421 GTCTTTGGGTTGGAGGGCTGGCTGCTCCAGGAGGCTGGCTGGCCGACCACAGCAACACCTCCCGGGCATC 480
           |||||| || || |||| ||||||||||||| |||| |||||||||| || || ||||||||
Db   421 AACACACCGGTCAACGGCTTCGAGGCCTTGGACCCGGCCTGGACCTGCCGC 480

Qy   481 AACATACCCGTGAATGGCTGATCCCCCGGTCCCCCGTGGGATCCGGAAGGCTGGGATCTGCCGC 540
           ||||||||| || |||||||||||| |||| |||||||||||| |||
Db   481 AACATACCCGTGAATGGCTGATCCCCCGGTCTCCCGTGGGATCCTGCCGC 540

Qy   541 CCGGCCCGCTTCAGCCTGCTCTCCTGCGTGGGCTGCCTGCGCGCCCTGGTGCACTCCGGGCTCTCCCAC 600
           |||||||||||||||||||| |||||||||||||||||
Db   541 CCTGCCCGACTCAGTTCTCGATTCGTCTTCCTTCCTTGTTTATCACTGTTTC 600

Qy   601 TGCTACGTGGGCTGCCTCCGGGCACTGGCGTCCGGCCCGTCCCCCCGCTCCGGACGGGCGAAGCTG 660
           ||||||||||||||||||||||||||||| || |||| |||||||||||||
Db   601 TGCTATGTGGGCTGCCTCCGGGCCCTGGTGCACTCCGGCCTGAGCCACAAACGGAAGCTC 660

Qy   661 CGGGCCGCCTGGGTGGCCGGGGGCCGGGGGCCCCTGCCTCGCTCGTCCTCGCTCCTACGAGACCCTAC 720
           |||||||||||||||| ||||| || ||| || | ||||||||| ||
Db   661 AGGGCCAGCTGGGTGGCTGGAGGAGCAGTTCTCACACTTCTCACACTCTCCTGGGCCCTAT 720

Qy   721 AACGCCCTCCAACGTGGCCAGCTTCCTGTACCCCAATCTAGGAGGCTCCTGGCGGAAGCTG 780
           ||||||||| || ||||| |||||||| ||||| ||| ||||||||| |||| || ||||
Db   721 AATGCTTCCAATGTGGCTAGTTTCATAAACCCGGACTTAGAGGGCTCCTGGAGGAAGTTG 780

Qy   781 GGGCTCATCACGGGGTGCTGGAGTGTGGTGCTTAATCCGGTTACTTGGGA 840
           ||| |||||| |||||||||||| ||||||||||||||||||||||| ||
Db   781 GGGCTCATCACAGGGGCGTGGAGTGTGGTGCTCAACCCACTGGTCACTTGGGA 840

Qy   841 AGGGGTCCTGGCCTGAAGACAGTGTGTGGCAAGAACCGAAGAGCCAAGTCCAGAAG 900
           |||| || ||||| ||| || || |||||||| | || |||| ||||||
Db   841 ACAGGTCCTGGACAGGGGGACAGGGGGACAGAGAGGACAATTCAGAAG 900

Qy   901 TA 902
           ||
Db   901 TA 902

, Mon May 3 09:26:58 2004

```
Qy   601 TGCTACGTGGGCTGCCTCCGGGCACTGGCCGCGCTCCGGCCTGACGCAGGCGGAAGCTG 660
             ||||| ||||||||||||||||||| |||| |||||||| ||||  ||||||||
Db   605 TGCTATGTGGGCTGCCTCCCGGGCCTCCGGTGCACTCGGGCCTGAGCCACAAACGGAAGCTC 664

Qy   661 CGGGCCGCCTGGGTGGCCGCGCGGGGCCCTCCTCACGCTGCTGCGTAGGACCCTAC 720
             |||||| ||||||||| |||| || | |||| ||| ||| ||||| | ||||||
Db   665 AGGGCAGCTTGGGTGGCTGCTGGAGGAGCACTTCTCACACTCCTGCCTCTGGGGCCCTAT 724

Qy   721 AACGCCTCCAACGTGGCCAGCTTCCTGTACCCCAATCTAGGAGGCTCCTGGCGGAAGCTG 780
             || |||||| |||| |||| |   ||||    |||| | |||||||||| ||| ||
Db   725 AATGCTTCCAATGTGGCTAGTTCATAAACCCGGACTTAGAAGGCTCCTGGAGGAAGTTG 784

Qy   781 GGGCTCATCACGGTGCCTGGAGTGTGGTGCTTAATCCGCTGGTGACCGGTTACTTGGGA 840
             |||||||||||||||||| |||||||| ||| | |||||| |||| ||||| ||||||
Db   785 GGGCTCATCACAGGAGCCTGGAGTGTGGTGCTCAACCCACTGGTCACTGGCTACTTGGGA 844

Qy   841 AGGGGTCCTGGCCTGAAGACAGTGTGCGGCAAGAACGCAAGGGGCAAGTCCCAGAAG 900
             | ||||||| |||| |  |  || ||| |||||| | ||| | ||||||| ||||||
Db   845 ACAGGTCCTGGACAGGGGACAATATGTGTGACCAGAGGGGACAATTCAGAAG 904

Qy   901 TA 902
             ||
Db   905 TA 906
```

Search completed: April 30, 2004, 17:38:31
Job time : 3931 secs

GenCore version 5.1.6

Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2004, 07:46:59 ; Search time 452 Seconds
(without alignments)
8486.997 Million cell updates/sec

Title: US-10-202-687-1
Perfect score: 903
Sequence: 903 1 atggacctgccccgcagct..........ggggcaagtcccagaagtaa 903

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04.*
1: geneseqn19808:*
2: geneseqn19908:*
3: geneseqn20008:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn20028:*
7: geneseqn2003aas:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn20048:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Query Length	DB	ID	Description
1	903	100.0	903	3	AAA30759	Aaa30759 Human G p
2	903	100.0	903	6	ABK90236	Abk90236 Human cDN
3	903	100.0	903	9	ABZ42843	Abz42843 Human G p
4	903	100.0	903	9	ADC22768	Adc22768 Human G p
5	900	99.7	900	9	ADB61405	Adb61405 DNA encod
6	898.2	99.5	903	9	AAA30776	Aaa30776 DNA encod
7	898.2	99.5	903	9	ADC22790	Adc22790 Human G-p
8	840.8	93.1	900	9	ADB61417	Adb61417 DNA encod
9	612	67.8	903	9	ADB61429	Adb61429 DNA encod
10	604.4	66.9	903	6	ABK90237	Abk90237 Mouse cDN
11	602.4	66.7	900	9	ADB61401	Adb61401 DNA encod
12	588	65.1	900	9	ADB61403	Adb61403 DNA encod
13	123.4	13.7	1841	2	AAV53631	Aav53631 Human 7-t
14	121.8	13.5	1041	3	AAA30762	Aaa30762 Human G p
15	121.8	13.5	1041	4	AAH26460	Aah26460 Human G-p
16	121.8	13.5	1041	4	AAH26459	Aah26459 Human G-p
17	121.8	13.5	1041	7	ABZ42561	Abz42561 Human G p
18	121.8	13.5	1041	9	ADC22772	Adc22772 Human G p
19	118.6	13.1	1041	9	AAA30777	Aaa30777 DNA encod
20	118.6	13.1	1041	9	ADC22792	Adc22792 Human G p
21	118.6	13.1	1041	9	AAX90981	Aax90981 Human pro
22	84.6	9.4	1534	3	AAA35309	Aaa35309 Human ade
23	84.6	9.4	1534	3	AAF21431	Aaf21431 Human low
24	84.6	9.4	1534	7	ABZ97125	Abz97125 Human nuc
25	84.6	9.4	4895	2	AAZ32747	Aaz32747 Human pro
26	84.6	9.4	4895	5	AAZ50775	Aaz50775 Human pro
27	84.6	9.4	4895	7	ABZ42864	Abz42864 Human pro
28	84.6	9.4	4925	4	AAH43632	Aah43632 Human PAR
29	84.6	9.4	6546	3	AAA35311	Aaa35311 Human ade
30	84.6	9.4	6546	3	AAF21433	Aaf21433 Human low
31	84.6	8.9	6546	7	ABZ97127	Abz97127 Human nuc
32	80	8.9	1098	3	AAA30740	Aaa30740 DNA encod
33	80	8.9	1098	3	AAA30666	Aaa30666 Human G p
34	80	8.9	1098	6	ABK48452	Abk48452 Human cDN
35	80	8.9	1098	7	AAD50860	Aad50860 Human G-p
36	80	8.9	1098	7	ABZ42746	Abz42746 Human G-p
37	80	8.9	1098	9	ADC22744	Adc22744 Human G p
38	80	8.9	1098	9	ADC22636	Adc22636 Human G p
39	80	8.9	1281	5	AAS77014	Aas77014 DNA encod
40	80	8.9	1281	5	AAS72718	Aas72718 DNA encod
41	80	8.9	1281	5	AAS79030	Aas79030 DNA encod
42	80	8.9	1431	5	AAS72235	Aas72235 DNA encod
43	80	8.9	1597	2	AAV07250	Aav07250 DNA encod
c 44	80	8.9	1629	5	AAS77311	Aas77311 DNA encod
45	80	8.9	1666	5	AAS92500	Aas92500 DNA encod

ALIGNMENTS

RESULT 1
AAA30759
ID   AAA30759 standard; cDNA; 903 BP.
XX
AC   AAA30759;
XX
DT   21-AUG-2000 (first entry)
XX
DE   Human G protein-coupled receptor GPR40 cDNA.
XX
KW   G protein-coupled receptor; GPCR; constitutively active;
KW   intracellular loop 3; transmembrane domain 6; drug screening; agonist;
KW   antagonist; ss.
XX
OS   Homo sapiens.
XX
PN   WO200022129-A1.
XX
PD   20-APR-2000.
XX
PF   12-OCT-1999; 99WO-US023938.
XX
PR   13-OCT-1998; 98US-00170496.
XX
PA   (AREN-) ARENA PHARM INC.
XX
PI   Behan DP, Chalmers DT, Liaw CW;
XX
DR   WPI; 2000-329165/28.
DR   P-PSDB; AAY90679.
XX
PT   Non-endogenous constitutively activated human G protein-coupled
PT   receptors, useful for identifying agonists for use as pharmaceutical
PT   agents.
XX
PS   Example 1; Page 307; 341pp; English.
XX
CC   The invention relates to constitutively active, non-endogenous versions
CC   of endogenous human orphan G protein-coupled receptors (GPCRs, AAY90643-
CC   AAY90677 and AAY90683-Y90687), and to DNA encoding them (AAA30709-A30743
CC   and AAA30775-A30779). The mutant proteins of the invention contain a
CC   mutation in a portion of the protein comprising intracellular loop 3
CC   (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, X,
CC   is substituted for an endogenous residue in IC3 at a position 16 amino
CC   acids N-terminal of an endogenous proline in TM6 to form a sequence X-
CC   (AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg or

CC   Ala, and is preferably Lys. When the endogenous residue at this position
CC   is Lys, this residue is replaced by His, Arg or preferably Ala. The 15
CC   amino acid stretch between the substituted amino acid and the Pro may be
CC   endogenous, non-endogenous, or a mixture of endogenous and non-endogenous
CC   residues. The constitutively active GPCRs are useful for identifying
CC   antagonists, agonists and partial agonists for use as pharmaceutical
CC   agents. The mutant proteins are also useful in research settings for
CC   elucidating the roles of the receptors in normal and diseased conditions.
CC   Antagonists for a particular GPCR are useful for treating diseases and
CC   disorders associated with that receptor. Because the novel mutant GPCRs
CC   are constitutively active, they can be used directly for screening of
CC   compounds without the need for endogenous ligands. The present sequence
CC   represents cDNA encoding a human wild-type GPCR used in an
CC   exemplification of the invention. This was cloned and subjected to site-
CC   directed mutagenesis (SDM) to generate DNA encoding the corresponding
CC   mutant of the invention
XX
SQ   Sequence 903 BP; 108 A; 327 C; 283 G; 185 T; 0 U; 0 Other;

Query Match   100.0%; Score 903; DB 3;   Length 903;
Best Local Similarity   100.0%;
  Matches 903; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 ATGGACCTGCCCCCGCCAGCTCTCCTTCGGCTCTATGTGGCCCGGCCTTTGCGCTGGGCTTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGGACCTGCCCCCGCCAGCTCTCCTTCGGCTCTATGTGGCCCGGCCTTTGCGCTGGGCTTC 60

Qy   61 CCGCTCAAGTCCTGGCCATCCGAGGGCGGCGACGCGCCAGCCCGCTCCGTCTCACCCCT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CCGCTCAAGTCCTGGCCATCCGAGGGCGGCGACGCGCCAGCCCGCTCCGTCTCACCCCT 120

Qy   121 AGCCTGGTCTACGCCCTGAACCTGGGCGTGCTCCGACCTGCTGACAGTCTCTCTGCCC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AGCCTGGTCTACGCCCTGAACCTGGGCGTGCTCCGACCTGCTGACAGTCTCTCTGCCC 180

Qy   181 CTGAAGGCGGTGGAGGCGTAGCCTCCGGGGCCTGGCCTTCGCCGGCCTCGCTGTGCCCC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CTGAAGGCGGTGGAGGCGTAGCCTCCGGGGCCTGGCCTTCGCCGGCCTCGCTGTGCCCC 240

Qy   241 GTCTTCGCGGTGGCCCACTTCTTCCCACTCTATGCGGGCGGGGGGGGTTCCTGGCCCCTG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GTCTTCGCGGTGGCCCACTTCTTCCCACTCTATGCGGGCGGGGGGGGTTCCTGGCCCCTG 300

Qy   301 AGTGCAGGCCGCTACCTGGGAGCAGCCTTCCCCCTGGGCTACCAAGCCTTCCGGAGGCG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 AGTGCAGGCCGCTACCTGGGAGCAGCCTTCCCCCTGGGCTACCAAGCCTTCCGGAGGCG 360

Qy   361 TGCTATTCCTGGGGGGTGTGCGCGGCCCATCTGGGCCTCCTGTGTCACCTGGGTCTG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TGCTATTCCTGGGGGGTGTGCGCGGCCCATCTGGGCCTCCTGTGTCACCTGGGTCTG 420

Qy   421 GTCTTTGGGTTGGAGGCTCCAGGAGGCTGGCTGACCACAGCAACACCTCCCTGGGCATC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GTCTTTGGGTTGGAGGCTCCAGGAGGCTGGCTGACCACAGCAACACCTCCCTGGGCATC 480

Qy   481 AACACACCGGTCAACGGCTCTCCGGTCTCCGCTGCCTGGGAGCCTGGAGCCTGCCGGC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 AACACACCGGTCAACGGCTCTCCGGTCTCCGCTGCCTGGGAGCCTGGAGCCTGCCGGC 540

Qy   541 CCGGCCCGCTCAGCCTTCTCTCTCCTCCTGCTCTTTTTCTGCCTCTTTTTGGCCATCAGCCCTTC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 CCGGCCCGCTCAGCCTTCTCTCTCCTCCTGCTCTTTTTCTGCCTCTTTTTGGCCATCAGCCCTTC 600

Qy   601 TGCTACGTGGCTGCCTCCGGCACTGGCCCGCTCCGGCCTGAGCGACGCACAGGCGGAAGCTG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 TGCTACGTGGCTGCCTCCGGCACTGGCCCGCTCCGGCCTGAGCGACGCACAGGCGGAAGCTG 660

Qy   661 CCGGCCCGCCTGGGTGGCCGCGGGCGCCCCTCACGCGTGCTGCCTAGCGTGCTGCCTAGGACCCTAC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 CGGGCCGCCTGGGGTGCCGCGGCCGCCCCTCCTCACGCTGCTCTGCGTGCTGCCTAGGACCCTTAC 720

Qy   721 AACGGCCTCCAACGTGGCCAGCTTCCTGTACCCCGCCAATCTAGGAGGCTCCTGCGCGGAAGCTG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   721 AACGGCCTCCAACGTGGCCAGCTTCCTGTACCCCCAATCTAGGAGGCTCCTGCGGGAAGCTG 780

Qy   781 GGGCTCATCACGGGTGCCTGGAGTGGTGCTTAATCCGCTGGTGACCGGTTACTTGGGA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 GGGCTCATCACGGGTGCCTGGAGTGGTGCTTAATCCGCTGGTGACCGGTTACTTGGGA 840

Qy   841 AGGGGTCCTGCCTGAAGACAGTGTGCGCCAAGAACGCAAGGGGGCAAGTCCCAGAAG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 AGGGGTCCTGCCTGAAGACAGTGTGCGGCAAGAACGCAAGGGGGCAAGTCCCAGAAG 900

Qy   901 TAA 903
          |||
Db   901 TAA 903

RESULT 2
ABK90236
ID   ABK90236 standard; cDNA; 903 BP.
XX
AC   ABK90236;
XX
DT   21-OCT-2002 (first entry)
XX
DE   Human cDNA encoding G protein-coupled receptor GPR40.
XX
KW   Human; ss; gene; GPR40; G protein-coupled receptor; type 2 diabetes;
KW   obesity; antidiabetic; neuroprotective; anorectic; cerebroprotective; Gq;
KW   G protein; reporter gene; glucose intolerance; insulin intolerance;
KW   neurodegenerative disease; Alzheimer's disease; stroke.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..903
FT   /*tag= a
FT   /product= "GPR40"
XX
PN   WO20025783-A2.
XX
PD   25-JUL-2002.
XX
PF   18-DEC-2001; 2001WO-US048985.
XX
PR   22-DEC-2000; 2000GB-00031527.
XX
PA   (GLAX) GLAXO GROUP LTD.
PA   (SMIK) SMITHKLINE BEECHAM PLC.
XX
PI   Andrews JL, Briscoe CP, Ignar DM, Muir AI, Sauls HR, Tadayyon M;
XX
DR   WPI; 2002-599726/64.
DR   P-PSDB; ABG31106.
XX
PT   Identifying GPR40 receptor ligand for treating disorders e.g. obesity,
PT   comprises detecting whether the test compound competitively inhibits the
PT   binding of a fatty acid GPR40 ligand to a GPR40 receptor.
XX
PS   Disclosure; Page 44-48; 53pp; English.
XX
CC   The invention relates to screening a test compound to determine whether
CC   the compound is a GPR40 receptor (G protein-coupled receptor) ligand
CC   comprises detecting whether the test compound competitively inhibits the
CC   binding of a fatty acid GPR40 ligand to a GPR40 receptor. Also included
CC   are (1) a method of screening a compound for GPR40 antagonist activity,
CC   comprising: (a) measuring any detectable signal produced by a reporter
CC   gene (comprising a reporter gene under the control of G protein Gq
CC   responsive transcriptional element); (b) detecting a decrease in reporter
CC   gene expression in the presence of both test compound and agonist; or (c)
CC   detecting whether the test compound decreases glucose-stimulated insulin
CC   release from mammalian pancreatic beta cells in the presence of a GPR40
CC   agonist, compared to glucose-stimulated insulin release that would occur
CC   due to the presence of the GPR40 agonist; (2) a method of screening a
CC   compound for GPR40 agonist activity, which comprises: (a) detecting any

CC   reporter gene expression; or (b) detecting whether the compound binds to
CC   GPR40 and increases glucose-stimulated insulin release from mammalian
CC   pancreatic beta cells. The method is useful for identifying GPR40
CC   antagonist or agonist compounds for treating disorders e.g. type 2
CC   diabetes, obesity, glucose or insulin intolerance, neurodegenerative
CC   disease (e.g. Alzheimer's disease) or stroke. The present sequence is the
XX   human cDNA for GPR40
SQ   Sequence 903 BP; 108 A; 327 C; 283 G; 185 T; 0 U; 0 Other;

Query Match   100.0%; Score 903; DB 6; Length 903;
Best Local Similarity   100.0%; Pred. No. 2e-167;
Matches 903; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 ATGGACCTGCCCCCGCCAGCTCTCCTTCGGCCTCTATGTGCCGCCTTGGCTGGCTTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGGACCTGCCCCCGCCAGCTCTCCTTCGGCCTCTATGTGCCGCCTTGGCTGGCTTC 60

Qy   61 CCGCTCAACGTCCTGGCCATCCGAGGCGGACGAGCCCACGCCCGGCTCCGTCTCACCCCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CCGCTCAACGTCCTGGCCATCCGAGGCGGACGAGCCCACGCCCGGCTCCGTCTCACCCCT 120

Qy   121 AGCCTGGTCTACGGCCCTGAACCTGGGCTGTCTCCGACCTGCTGTGACAGTCTCTGCCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AGCCTGGTCTACGGCCCTGAACCTGGGCTGTCTCCGACCTGCTGTGACAGTCTCTGCCC 180

Qy   181 CTGAAGGCGGTGGAGGCGGTGAGGCGGTGAGCCTCCGGGGCCTGGCCTCGCGCCTGTGCCCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CTGAAGGCGGTGGAGGCGGTGAGGCGGTGAGCCTCCGGGGCCTGGCCTCGCGCCTGTGCCCC 240

Qy   241 GTCTTCGCGGTGCGCCACTTCTTCCCACTCTATGCCGCGGCGGGGGGCTTCCTGGCCCGCCCCTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GTCTTCGCGGTGCGCCACTTCTTCCCACTCTATGCCGCGGCGGGGGGCTTCCTGGCGCGCCCCTG 300

Qy   301 AGTGCAGGCCGCTACCTGGGAGCAGCCTTCCCCCTTGGGCTACCAAGCCTTCCGGAGGCCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 AGTGCAGGCCGCTACCTGGGAGCAGCCTTCCCCCTTGGGCTACCAAGCCTTCCGGAGGCCG 360

Qy   361 TGCTATTCCTGGGGGTGTGCGCGGCCATCTGGGCCTTCGTCCTGTCTCCTGGTCTTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TGCTATTCCTGGGGGTGTGCGCGGCCATCTGGGCCTTCGTCCTGTCTCCTGGTCTTG 420

Qy   421 GTCTTTGGGTTGGAGGCTTCCAGGAGGCTGGCTGGACCAGCAACACCTCCCTGGGCATC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GTCTTTGGGTTGGAGGCTTCCAGGAGGCTGGCTGGACCAGCAACACCTCCCTGGGCATC 480

Qy   481 AACACACCGGTCAACGGCTTCTCCGGTCTGCCTGGAGGCCGGACCCGGCCTCTGCCGGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 AACACACCGGTCAACGGCTTCTCCGGTCTGCCTGGAGGCCGGACCCGGCCTCTGCCGGC 540

Qy   541 CCGGCCCGGCTCAGCCTTCCCGCTCTTTTTCTGCCCCTTGGCCATCACAGCCTTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 CCGGCCCGGCTTCAGCCTTCCCGCTCTTTTTCTGCCCCTTGGCCATCACAGCCTTC 600

Qy   601 TGCTACGTGGGGCTGCCTCCGGCACTGGCCCGTCCGGCCCGTGACGCACAGGCGGAAGCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 TGCTACGTGGGGCTGCCTCCGGCACTGGCCCGTCCGGCCCGTGACGCACAGGCGGAAGCTG 660

Qy   661 CCGGCCGCCTGGGTGGCCGGCGGGGGCCCCTCCACGCTGCTCTGCGTAGGACCCTAC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 CGGGCCGCCTGGGTGGCCGGCGGGGGCCCCTCCACGCTGCTCTGCGTGCGTAGGACCCTAC 720

Qy   721 AACGCCTTCCAACGTGGGCCAGCTTCCTGTACCCCAATCTAGGAGGCTCCTGGCCGGAAGCTG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 AACGCCTTCCAACGTGGCCAGCTTCCTGTACCCCAATCTAGGAGGCTCCTGGCCGGAAGCTG 780

Qy   781 GGGCTCATCACGGGGTGCCTGGAGTGTGGTTGCTTAATCCGCTGGTTAATCCGGTTACTTGGGA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 GGGCTCATCACGGGGTGCCTGGAGTGTGGTTGCTTAATCCGCTGGTTAATCCGGTTACTTGGA 840

Qy   841 AGGGGTCCTGGCCTGAAGACACAGTGTGCGGCAAGAACGCAAGGGGCAAGTCCCAGAAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 AGGGGTCCTGGCCTGAAGACACAGTGTGCGGCAACAACGCAAGGGGCAAGTCCCAGAGAG 900

Qy   901 TAA 903
          |||
Db   901 TAA 903

RESULT 3
ABZ42843
ID   ABZ42843 standard; DNA; 903 BP.
XX
AC   ABZ42843;
XX
DT   04-MAR-2003 (first entry)
XX
DE   Human G protein-coupled receptor GPR40 nucleotide SEQ ID NO:474.
XX
KW   G protein-coupled receptor; GPCR; antigenic peptide; gene therapy;
KW   G protein-coupled receptor modulator; antibody; immune-related disease;
KW   growth-related disease; cell regeneration-related disease; AIDS; cancer;
KW   immunological-related cell proliferative disease; autoimmune disease;
KW   Alzheimer's disease; atherosclerosis; infection; osteoarthritis; allergy;
KW   osteoporosis; cardiomyopathy; inflammation; Crohn's disease; diabetes;
KW   graft versus host disease; Parkinson's disease; multiple sclerosis; pain;
KW   psoriasis; anxiety; depression; schizophrenia; dementia; memory loss;
KW   mental retardation; epilepsy; asthma; tuberculosis; obesity; nausea;
KW   hypertension; hypotension; renal disorder; rheumatoid arthritis; trauma;
KW   ulcer; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200261087-A2.
XX
PD   08-AUG-2002.
XX
PF   19-DEC-2001; 2001WO-US050107.
XX
PR   19-DEC-2000; 2000US-0257144P.
XX
PA   (LIFE-) LIFESPAN BIOSCIENCES INC.
XX
PI   Burmer GC, Roush CL, Brown JP;
XX
DR   WPI; 2003-046718/04.
DR   P-PSDB; ABP81994.
XX
PT   New isolated antigenic peptides e.g., for G protein-coupled receptors
PT   (GPCR), useful for diagnosing and designing drugs for treating conditions
PT   in which GPCRs are involved, e.g. AIDS, Alzheimer's disease, cancer or
PT   autoimmune diseases.
XX
PS   Disclosure; Fig 1; 523pp; English.
XX
CC   The present invention describes antigenic peptides (I) comprising: (a)
CC   any one of 1601 sequences (see ABPB2019 to ABP83619) of 12-24 amino
CC   acids. Also described: (1) an assay for the detection of a particular G
CC   protein-coupled receptor (GPCR) or a candidate polypeptide in a sample;
CC   and (2) an isolated antibody having high specificity and high affinity or
CC   avidity for a particular GPCR. (I) can be used as GPCR modulators and in
CC   gene therapy. The antigenic peptides for GPCRs are useful in detecting an
CC   antibody against a particular GPCR, and in the production of specific
CC   antibodies. The peptides and antibodies are also useful for detecting the
CC   presence or absence of corresponding GPCRs. The antigenic peptides for
CC   GPCRs and antibodies are useful for diagnosing and designing drugs for
CC   treating immune-related diseases, growth-related diseases, cell
CC   regeneration-related disease, immunological-related cell proliferative
CC   diseases or autoimmune diseases, e.g. AIDS, Alzheimer's disease,
CC   atherosclerosis, bacterial, fungal, protozoan or viral infections,
CC   osteoarthritis, osteoporosis, cancer, cardiomyopathy, chronic and acute
CC   inflammation, allergies Crohn's disease, diabetes, graft versus host
CC   disease, Parkinson's disease, multiple sclerosis, pain, psoriasis,
CC   anxiety, depression, schizophrenia, dementia, mental retardation, memory
CC   loss, epilepsy, asthma, tuberculosis, obesity, nausea, hypertension,
CC   hypotension, renal disorders, rheumatoid arthritis, trauma, ulcers, or

CC   any other disorder in which GPCRs are involved. The antibodies may be
CC   used in immunoassays and immunodiagnosis. ABZ42523 to ABZ42869 encode
CC   GPCR proteins given in ABP81675 to ABP82018, which are used in the
CC   exemplification of the present invention
XX
SQ   Sequence 903 BP; 108 A; 327 C; 283 G; 185 T; 0 U; 0 Other;

Query Match   100.0%; Score 903; DB 7; Length 903;
Best Local Similarity   100.0%; Pred. No. 2e-167;
Matches 903; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 ATGGACCTGCCCCCGCCAGCTCTCCTTCGGCCTCCTATGTGGCCGCCTTTGCGCTGGGCTTC 60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGGACCTGCCCCCGCCAGCTCTCCTTCGGCCTCCTATGTGGCCGCCTTTGCGCTGGGCTTC 60

Qy   61 CCGCTCAACGTCCTGGCCATCCAGAGGCGGACGGCCCACGCCCCGCTCGGTCTCACCCCT 120
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CCGCTCAACGTCCTGGCCATCCAGAGGCGGACGGCCCACGCCCCGCTCGGTCTCACCCCT 120

Qy   121 AGCCTGGTCTACGGCCCTGAACCTGGGCTGCTCCGACCTGCTGACAGTCTCTCTGCCC 180
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AGCCTGGTCTACGGCCCTGAACCTGGGCTGCTCCGACCTGCTGACAGTCTCTCTGCCC 180

Qy   181 CTGAAGGCGGTGAGGCGCTAGCCTCCGGGGCTGGCCTTCGCCTCGTGTGCCC 240
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CTGAAGGCGGTGAGGCGCTAGCCTCCGGGGCTGGCCTTCGCCTCGTGTGCCC 240

Qy   241 GTCTTCGCGGTGGCCCACTTCTTCCCACTCTATGCCGGGGGGCTTCCTGGCCGCCCTG 300
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GTCTTCGCGGTGGCCCACTTCTTCCCACTCTATGCCGGGGGGCTTCCTGGCCGCCCTG 300

Qy   301 AGTGCAGGCCGCTACCTGGGAGCAGCCTTCCCCCTGGGCTACCAAGCTTCCGGAGGCCG 360
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 AGTGCAGGCCGCTACCTGGGAGCAGCCTTCCCCCTGGGCTACCAAGCTTCCGGAGGCCG 360

Qy   361 TGCTATTCCTGGGGGTGCGCGGCCATCTGGGCCCCATCTGTCTGTGTCACCTGGGTCTG 420
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TGCTATTCCTGGGGGTGCGCGGCCATCTGGGCCCCATCTGTCTGTGTCACCTGGGTCTG 420

Qy   421 GTCTTGGGTTGGAGGCTCCAGGAGGCTGCTCGACCACAGCAACACCTCCCTGGCATC 480
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GTCTTGGGTTGGAGGCTCCAGGAGGCTGCTCGACCACAGCAACACCTCCCTGGCATC 480

Qy   481 AACACCACCGGTCAACGGCTCTCCGGTCTGCCTGGAGGGCTGGACCCGGCCTCTGCCGGC 540
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 AACACCACCGGTCAACGGCTCTCCGGTCTGCCTGGAGGGCTGGACCCGGCCTCTGCCGGC 540

Qy   541 CCGGCCCGCTTCAGCCTTCTCTCTGCTCTTTTTCTGCCCTTGGCCATCACAGCCTTC 600
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 CCGGCCCGCTTCAGCCTTCTCTCTGCTCTTTTTCTGCCCTTGGCCATCACAGCCTTC 600

Qy   601 TGCTACGTGGGCTGCCTCCGGGCACTGGCCCGCTCGACGCAGCACAGGCCGGAAGCTG 660
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 TGCTACGTGGGCTGCCTCCGGGCACTGGCCCGCTCGACGCAGCACAGGCCGGAAGCTG 660

Qy   661 CCGGGCCGCTGGGTGGCCGGGGCCCCTCCTACGCTGCGTCTGCGGCTAGGACCCTAC 720
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 CGGGGCCGCTGGGGTGGCCGGGGCCCCTCCTACGCTGCGTCTGCGGCTAGGACCCCTAC 720

Qy   721 AACGCCCTCCAAGTGGCCAGCTTCCTGTACCCCAATCTAGGAGGCTCCTGGCCGAAGCTG 780
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 AACGCCCTCCAAGTGGCCAGCTTCCTGTACCCCAATCTAGGAGGCTCCTGGCCGGAAGCTG 780

Qy   781 GGGCTCATCACGGGTGCCTGGAGTGTGGTGCTTAATCCCGTGGTTACTTGGGA 840
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 GGGCTCATCACGGGTGCCTGGAGTGTGGTGCTTAATCCCGTGGTTACTTGGGA 840

Qy   841 AGGGGTCCTGGCCTGAAAGACAGTGTGTGCGGCAAGAACCGCAAGGGGGCAAGTCCCAGAAG 900
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 AGGGGTCCTGGCCTGAAAGACAGTGTGTGCGGCAAGAACCGCAAGGGGGCAAGTCCCAGAAG 900

Qy   901 TAA 903
             |||
Db   901 TAA 903

Db   901 TAA 903

RESULT 4
ID   ADC22768   standard; cDNA; 903 BP.
XX
AC   ADC22768;
XX
DT   18-DEC-2003 (first entry)
XX
DE   Human G protein-coupled receptor cDNA #36.
XX
KW   Human; gene; ss; G protein-coupled receptor; GPCR;
KW   transmembrane-6 region; TM6; intracellular-3 region; IC3.
XX
OS   Homo sapiens.
XX
PN   US6555339-B1.
XX
PD   29-APR-2003.
XX
PF   13-OCT-1998;   98US-00170496.
XX
PR   14-APR-1997;   97US-00839449.
PR   14-APR-1998;   98US-00060188.
PR   26-JUN-1998;   98US-0090783P.
PR   07-AUG-1998;   98US-0095677P.
XX
PA   (AREN-) ARENA PHARM INC.
XX
PI   Liaw CW, Behan DP, Chalmers DT;
XX
DR   WPI; 2003-742861/70.
DR   P-PSDB; ADC22769.
XX
PT   Creating a constitutively active version of an endogenous human G protein
PT   coupled receptor (GPCR) comprises substituting a specific amino acid in
PT   the transmembrane-6 region with a different amino acid, and testing for
PT   constitutive activity.
XX
PS   Example 1; SEQ ID NO 249; 221pp; English.
XX
CC   The invention relates to a method for treating a non-endogenous,
CC   constitutively active version of an endogenous human G protein-coupled
CC   receptor (GPCR) that has a transmembrane-6 (TM6) region and an
CC   intracellular-3 (IC3) region, by substituting a specific amino acid in
CC   the TM6 region with a different amino acid, and testing for constitutive
CC   activity. The method is useful for creating a constitutively active
CC   version of an endogenous human GPCR that comprises a transmembrane 6
CC   region and an intracellular loop 3 region. The altered human GPCR
CC   polypeptides are useful for screening test compounds for identification
CC   of inverse agonists or partial agonists of GPCR polypeptides, which may
CC   have therapeutic uses. The altered GPCRs may also be used in vivo or in
CC   vitro in biological research. A nucleic acid encoding the altered GPCR
CC   may be used to create a transgenic animal expressing the altered GPCR.
CC   The method allows screening for compounds that modulate the activity of a
CC   human G protein-coupled receptor without the need for provision of a
CC   ligand for the receptor. This is particularly useful in allowing
CC   screening of compounds against orphan receptors for which no ligand is
CC   currently known. This sequence represents cDNA encoding a human GPCR
CC   polypeptide of the invention.
XX
SQ   Sequence 903 BP; 108 A; 327 C; 283 G; 185 T; 0 U; 0 Other;

Query Match   100.0%; Score 903; DB 9; Length 903;
Best Local Similarity   100.0%; Pred. No. 2e-167;
Matches 903; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 ATGGACCTGCCCCCGCCAGCTCTCCTTCGGCCTCCTATGTGGCGCCTTTGCGCTGGGCTTC 60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGGACCTGCCCCCGCCAGCTCTCCTTCGGCCTCCTATGTGGCGCCTTTGCGCTGGGCTTC 60

KW antilipaemic; dermatological; antiarteriosclerotic; antiarthritic;
KW osteopathic; thrombolytic; nootropic; anorectic; hypotensive; cytostatic;
KW diabetes; ketosis; acidosis; diabetic neuropathy; diabetic nephropathy;
KW diabetic retinopathy; hyperlipaemia; skin disease; arthritis;
KW bone disease; arteriosclerosis; thrombosis; hypertension; indigestion;
KW memory loss; obesity; hypoglycaemia; edema; insulin resistance;
KW insulin allergy; fat metabolism disorder; cancer; human; gene; ds.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT CDS   1..900
FT   /*tag= a
FT   /partial
FT   /note= "No stop codon"
FT   /product= "Human GPR40 protein"
XX
PN WO2003068959-A1.
XX
PD 21-AUG-2003.
XX
PF 13-FEB-2003; 2003WO-JP001483.
XX
PR 14-FEB-2002; 2002JP-00037131.
PR 12-JUL-2002; 2002JP-00204163.
PR 12-NOV-2002; 2002JP-00328696.
PR 22-JAN-2003; 2003JP-00014032.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Hinuma S, Hosoya M, Ito Y, Kobayashi M, Tanaka H, Okubo S;
PI Fujii R, Kizawa H, Kawamata Y, Ogi K, Harada M, Fukusumi S;
XX
DR WPI; 2003-671661/63.
DR P-PSDB; ADB61404.
XX
PT Fatty acid and eicosanoid-binding G-protein coupled receptor protein
PT GPR40 for control of pancreatic function and treatment of diabetes.
XX
PS Disclosure; Page 235-237; 257pp; Japanese.
XX
CC The invention relates to a novel screening method comprising a fatty acid
CC and eicosanoid-binding G-protein coupled receptor protein, GPR40,
CC originating in mouse, rat, crab-eating monkey and hamster (sequences
CC fully defined in the specification), equivalent proteins of similar
CC activity, and peptides containing partial sequences of the GPR40 protein.
CC The novel fatty acid and GPR40 protein and their compositions have the
CC following activities: antidiabetic, anabolic, neuroprotective,
CC nephrotropic, ophthalmological, antilipaemic, dermatological,
CC antiarteriosclerotic, antiarthritic, osteopathic, thrombolytic,
CC nootropic, anorectic, hypotensive, and cytostatic. The novel fatty acid
CC and eicosanoid-binding G-protein coupled receptor protein, GPR40, and
CC further compositions can be used in the treatment, prevention and
CC diagnosis of diabetes, ketosis, acidosis, diabetic neuropathy, diabetic
CC nephropathy, diabetic retinopathy, hyperlipaemia, skin diseases,
CC arthritis, bone diseases, arteriosclerosis, thrombosis, hypertension,
CC indigestion, memory loss, obesity, hypoglycaemia, edema, insulin
CC resistance, insulin allergy, fat metabolism disorders and cancer. This
CC polynucleotide sequence represents the DNA encoding a human GPR40 protein
CC of the invention.
XX
SQ Sequence 900 BP; 106 A; 327 C; 283 G; 184 T; 0 U; 0 Other;

Query Match   99.7%; Score 900; DB 9; Length 900;
Best Local Similarity 100.0%; Pred. No. 7.8e-167;
Matches 900; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 ATGGACCTGCCCCCGCCAGCTCTCTTCGGCCTCTATGTGGCCGCCCTTTGCCGCTGGGCTTC 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGGACCTGCCCCCGCCAGCTCTCTTCGGCCTCTATGTGGCCGCCCTTTGCCGCTGGGCTTC 60

Qy   61 CGGCTCAACGTCCTGCCCATCGGAGGCGGCCACGCCCGGCTCCGTCTCCACCCCT 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy   61 CGGCTCAACGTCCTGCCCATCGGAGGCGGCCACGCCCGGCTCCGTCTCCACCCCT 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CGGCTCAACGTCCTGCCCATCGGAGGCGGCCACGCCCGGCTCCGTCTCCACCCCT 120

Qy   121 AGCGCTGGTCACGCCCTGAACCTGGGCTGCTCCGACCTGCTGCACAGTCTCTGCCC 180
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AGCGCTGGTCACGCCCTGAACCTGGGCTGCTCCGACCTGCTGCACAGTCTCTGCCC 180

Qy   181 CTGAAGGCGGTGGAGGCGCTAGCCTTCCGGGGCCTTGCCTTCGTGTGCCCC 240
             |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CTGAAGGCGGTGGAGGCGCTAGCCTTCCGGGGCCTTGCCTTCGTGTGCCCC 240

Qy   241 GTCTTCGCGGTGGCCCACTTCTTCCCACTCTATGCCGGCGGGGGCTTCTGGCGGCCCCTG 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GTCTTCGCGGTGGCCCACTTCTTCCCACTCTATGCCGGCGGGGGCTTCTGGCGGCCCCTG 300

Qy   301 AGTGCAGGCGCGCTACCTGGGAGCAGCAGCCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCG 360
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 AGTGCAGGCGCGCTACCTGGGAGCAGCAGCCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCG 360

Qy   361 TGCTATTCCTGGGGGGTGTGCGCGGCCATCTGGGCCCTCGTCCTGTGTCACCTGGGTCTG 420
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TGCTATTCCTGGGGGGTGTGCGCGGCCATCTGGGCCCTCGTCCTGTGTCACCTGGGTCTG 420

Qy   421 GTCTTGGGTTGGAGGCTCCAGGAGGCTGGCTGAGACCAACACCCTCCCTGGGCATC 480
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GTCTTTGGGTTGGAGGCTCCAGGAGGCTGGCTGAGACCAACACCCTCCCTGGGCATC 480

Qy   481 AACACCACCGTCAACGGCTTCTCCGGCTCTCGCCTGGAGGCGCCTCGTCTGCCGGC 540
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 AACACCACCGTCAACGGCTTCTCCGGCTCTCGCCTGGAGGCGCCTCGTCTGCCGGC 540

Qy   541 CCGGCCGCGCTTCAGCCTCTTCTCTCCGCTCTTTTTCTGCCCCTTGGCCATCACAGCCTTC 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 CCGGCCCGCGCTTCAGCCTCTTCTCTCCGCTCTTTTTCTGCCCCTTGGCCATCACAGCCTTC 600

Qy   601 TGCTACGTGGGCTGCTCCTCCGGCACTGCCCGCCGCTTCCGGCCTGACGGCACAGGCGGAAGCTG 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 TGCTACGTGGGCTGCTCCTCCGGCACTGCCCGCCGCTTCCGGCCTGACGGCACAGGCGGAAGCTG 660

Qy   661 CGGGCCGGCCTGGGTGGCCCGGCGGGGGCCCTCCACGCACGCGGCGCTCTGCGGTAGGACCCTAC 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 CGGGCCGGCCTGGGTGGCCCGGCGGGGGCCCCTCCACGCTGCTCCGGTAGGACCCTAC 720

Qy   721 AACGCCTCCAACGTGGCCAGCTTCCTGTACCCCAATCTAGGAGGCTCCTGGCGGAAGCTG 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 AACGCCTCCAACGTGGCCAGCTTCCTGTACCCCAATCTAGGAGGCTCCTGGCGGAAGCTG 780

Qy   781 GGGCTCATCACGGGTGCCTGGAGTGTGGTGCTTAATCCGCTGGTGACCCGGTTACTTGGGA 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 GGGCTCATCACGGGTGCCTGGAGTGTGGTGCTTAATCCGCTGGTGACCCGGTTACTTGGGA 840

Qy   841 AGGGGTCCTGGCCTGAAGAGACAGTGTGCGGCCAAGAACGCAAGGGGCAAGTCCCAGAAG 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 AGGGGTCCTGGCCTGAAGAGACAGTGTGCGGCCAAGAACGCAAGGGGCAAGTCCCAGAAG 900

Qy   901 TAA 903
             |||
Db   901 TAA 903

RESULT 5
ADB61405
ID ADB61405 standard; DNA; 900 BP.
XX
AC ADB61405;
XX
DT 04-DEC-2003 (first entry)
XX
DE DNA encoding human GPR40 protein.
XX
KW fatty acid; eicosanoid-binding G-protein coupled receptor; GPR40;
KW antidiabetic; anabolic; neuroprotective; nephrotropic; ophthalmological;

Db   61 CCGCTCAACGTCCTGGCCATCCGAGGCGGCGACGGGCCCACGCGCCTCCGTCTCCACCCCT 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   121 AGCCTGGTCTACGCCCTGAACCTGGGCTGCTGCTGCTCCGACCTGCTGACAGTCTCTCTGCCC 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AGCCTGGTCTACGCCCTGAACCTGGGCTGCTGCTGCTCCGACCTGCTGACAGTCTCTCTGCCC 180
Qy   181 CTGAAGGCGGTGGAGGGCGTAGCCTCCGGGGCCTCGCCGGCCCTCGCTCGTGCCCC 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CTGAAGGCGGTGGAGGGCGTAGCCTCCGGGGCCTCGCCGGCCCTCGCTCGTGCCCC 240
Qy   241 GTCTTCGCGGTGGCGTGGCCCACTTCTTCCCACTCTATGCCGGCGGGGGCTTCCTGGCGGCCCTG 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GTCTTCGCGGTGGCGTGGCCCACTTCTTCCCACTCTATGCCGGCGGGGGCTTCCTGGCGGCCCTG 300
Qy   301 AGTGCAGGGCGGCGCTACCTGGGAGCAGCCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCGCG 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 AGTGCAGGGCGGCGCTACCTGGGAGCAGCCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCGCG 360
Qy   361 TGCTATTCCTGGGGGTGTGCGCGCGCCATCTGGGCCCTCGTCCTGTCACCTGGGTCTG 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TGCTATTCCTGGGGGTGTGCGCGCGCCATCTGGGCCCTCGTCCTGTCACCTGGGTCTG 420
Qy   421 GTCTTTGGGTTGAGGCTCCAGGAGGCTGGCTGGAGGGCTGGACCACAGCAACACCTCCCTGGGCATC 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GTCTTTGGGTTGAGGCTCCAGGAGGCTGGCTGGAGGGCTGGACCACAGCAACACCTCCCTGGGCATC 480
Qy   481 AACACACCGGTCAACGGCTCTCCGGTCTGCCTGGAGGCTGGACCCGGGCCTCTGCCGGC 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 AACACACCGGTCAACGGCTCTCCGGTCTGCCTGGAGGCTGGACCCGGGCCTCTGCCGGC 540
Qy   541 CCGGCCCGCTTCAGCCTCTCTCTCTCTGCCTTTTTCTGGCCATCACAGCCTTC 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 CCGGCCCGCTTCAGCCTCTCTCTCTCTGCCTTTTTCTGGCCATCACAGCCTTC 600
Qy   601 TGCTACGTGGGCTGCCTCCGGGACTGGCCCGCTGCCGGCCGTGACGCACAGGCCGAAGCTG 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 TGCTACGTGGGCTGCCTCCGGGACTGGCCCGCTGCCGGCCGTGACGCACAGGCCGAAGCTG 660
Qy   661 CGGGCCGCCTGGGTGGCCGGCGGGGCCCCTCACGCGTGCCTGCTGCTGCTGCTGCTGGAGGCTCCTGGCCGGAAGTG 720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 CGGGCCGCCTGGGTGGCCGGCGGGGCCCCTCACGCGTGCCTGCTGCTGCTGCTGCTGGAGGCTCCTGGCCGGAAGTG 720
Qy   721 AACGGCCTCAACGTGGCCAGCTTCCTGTACCCCAATCTAGGAGGCTCCTGGCCGGAAGCTG 780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 AACGGCCTCAACGTGGCCAGCTTCCTGTACCCCAATCTAGGAGGCTCCTGGCCGGAAGCTG 780
Qy   781 GGGGTCATCACGGGGTGCCTGGAGTGTGGTGCTTAATCCGCTGGTGCTTAATCCGCTGGTTACTTGGGA 840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 GGGGTCATCACGGGGTGCCTGGAGTGTGGTGCTTAATCCGCTGGTGCTTAATCCGCTGGTTACTTGGGA 840
Qy   841 AGGGGTCCTGGCCTGAAGACAGTGTGCGGCAAGAACGCAAGGGGCAAGTCCCAGAAG 900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 AGGGGTCCTGGCCTGAAGACAGTGTGCGGCAAGAACGCAAGGGGCAAGTCCCAGAAG 900

RESULT 6
AAA30776
ID   AAA30776 standard; DNA; 903 BP.
XX
AC   AAA30776;
XX
DT   21-AUG-2000 (first entry)
XX
DE   DNA encoding human mutant G protein-coupled receptor GPR40 (A223K).
XX
KW   G protein-coupled receptor; GPCR; constitutively active;
KW   intracellular loop 3; transmembrane domain 6; drug screening; agonist;
KW   antagonist; mutant; ss.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200022129-A1.

PT   Non-endogenous constitutively activated human G protein-coupled
PT   receptors, useful for identifying agonists for use as pharmaceutical
PT   agents.
XX
PS   Example 2; Page 324; 341pp; English.
XX
CC   The invention relates to constitutively active, non-endogenous versions
CC   of endogenous human orphan G protein-coupled receptors (GPCRs, AAY90643-
CC   AAY90677 and AAY90683-Y90687). The mutant proteins of the invention contain a
CC   mutation in a portion of the protein comprising intracellular loop 3
CC   (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, X,
CC   is substituted for an endogenous residue in IC3 at a position 16 amino
CC   acids N-terminal of an endogenous proline in TM6 to form a sequence X-
CC   (AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg or
CC   Ala, and is preferably Lys. When the endogenous residue at this position
CC   is Lys, this residue is replaced by His, Arg or preferably Ala. The 15
CC   amino acid stretch between the substituted amino acid and the Pro may be
CC   endogenous, non-endogenous, or a mixture of endogenous and non-endogenous
CC   residues. The constitutively active GPCRs are useful for identifying
CC   antagonists, agonists and partial agonists for use as pharmaceutical
CC   agents. The mutant proteins are also useful in research settings for
CC   elucidating the roles of the receptors in normal and diseased conditions.
CC   Antagonists for a particular GPCR are useful for treating diseases and
CC   disorders associated with that receptor. Because the novel mutant GPCRs
CC   are constitutively active, they can be used directly for screening of
CC   compounds without the need for endogenous ligands. Sequences AAA30709-
CC   AAA30743 and AAA30775-A30779 represent DNAs encoding the mutant human
CC   GPCRs of the invention
XX
SQ   Sequence 903 BP; 111 A; 325 C; 282 G; 185 T; 0 U; 0 Other;

Query Match   99.5%;   Score 898.2;   DB 3;   Length 903;
Best Local Similarity   99.7%;   Pred. No. 1.7e-166;
Matches 900;   Conservative 0;   Mismatches 3;   Indels 0;   Gaps 0;

Qy   1 ATGGACCTGCCCCCGCCAGCTCCTTCTGGCCTCATGTGGCCTCGGTCTCCATGTGCCGGCTTC 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGGACCTGCCCCCGCCAGCTCCTTCTGGCCTCATGTGGCCTCGGTCTCCATGTGCCGGCTTC 60

Qy   61 CCGGCTCAACGTCCTGGCCATCCGAGGCGGCGACGGGCCCACGCGCCTCCGTCTCCACCCCT 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CCGGCTCAACGTCCTGGCCATCCGAGGCGGCGACGGGCCCACGCGCCTCCGTCTCCACCCCT 120

Qy 121 AGCCTGGTCTACGCCCTGAACCTGGGCTGCTGCTGCTGACAGTCTCTGACCAGTCTCTCTGCCC 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 AGCCTGGTCTACGCCCTGAACCTGGGCTGCTGCTGCTGACAGTCTCTGACCAGTCTCTCTGCCC 180

Qy 181 CTGAAGGCGGTGGAGGGCGTAGCCTGGCGCCTAGCCTGCCTGCCGTGTGCCCC 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 CTGAAGGCGGTGGAGGGCGTAGCCTGGCGCCTAGCCTGCCTGCCGTGTGCCCC 240

Qy 241 GTCTTCGCGGTGGCGTGGCCCACTTCTTCCCACTCTATGCCGGCGGGGGCTTCCTGGCCGCCCTG 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GTCTTCGCGGTGGCGTGGCCCACTTCTTCCCACTCTATGCCGGCGGGGGCTTCCTGGCCGCCCTG 300

Qy 301 AGTGCAGGGCCGCTACCTGGGGGAGCAGCAGCGCTACCAAGCCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCG 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 AGTGCAGGGCCGCTACCTGGGGGAGCAGCAGCGCTACCAAGCCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCG 360

DR   P-PSDB; ADC22791.
XX
PT   Creating a constitutively active version of an endogenous human G protein
PT   coupled receptor (GPCR) comprises substituting a specific amino acid in
PT   the transmembrane-6 region with a different amino acid, and testing for
PT   constitutive activity.
XX
PS   Example 2; SEQ ID NO 271; 221pp; English.
XX
CC   The invention relates to a method for treating a non-endogenous,
CC   constitutively active version of an endogenous human G protein-coupled
CC   receptor (GPCR) that has a transmembrane-6 (TM6) region and an
CC   intracellular-3 (IC3) region, by substituting a specific amino acid in
CC   the TM6 region with a different amino acid, and testing for constitutive
CC   activity. The method is useful for creating a constitutively active
CC   version of an endogenous human GPCR that comprises a transmembrane 6
CC   region and an intracellular loop 3 region. The altered human GPCR
CC   polypeptides are useful for screening test compounds for identification
CC   of inverse agonists or partial agonists of GPCR polypeptides, which may
CC   have therapeutic uses. The altered GPCRs may also be used in vivo or in
CC   vitro in biological research. A nucleic acid encoding the altered GPCR.
CC   may be used to create a transgenic animal expressing the altered GPCR.
CC   The method allows screening for compounds that modulate the activity of a
CC   human G protein-coupled receptor without the need for provision of a
CC   ligand for the receptor. This is particularly useful in allowing
CC   screening of compounds against orphan receptors for which no ligand is
CC   currently known. This sequence represents cDNA encoding a human GPCR
CC   polypeptide of the invention.
XX
SQ   Sequence 903 BP; 111 A; 325 C; 282 G; 185 T; 0 U; 0 Other;

Query Match   99.5%; Score 898.2; DB 9; Length 903;
Best Local Similarity   99.7%; Pred. No. 1.7e-166;
Matches 900; Conservative 0; Mismatches 3; Indels 0; Gaps .0;

Qy   1 ATGGACCTGCCCCCGAGCTCCTCTTCGGCCTCCTATGTGCGCGTGGCTTCGGCGTGGGGTTC 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGGACCTGCCCCCGAGCTCCTCTTCGGCCTCCTATGTGGCGCGTGGCTTCGGCGTGGGGTTC 60

Qy   61 CCGCTCAAGTCTCTGCCATCCGAGGCGCGACGGCCCAGCCCGGCTCGTCCGGCTCGGGGCCCCT 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CCGCTCAAGTCTCTGCCATCCGAGGCGCGACGGCCCAGCCCGGCTCGTCCGGCTCGGGGCCCCT 120

Qy 121 AGCCTGGTCTACGCCCGGAACCTGGCTGCTGTCCGACCTGCTGCTGACAGTCTCTGCCCC 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 AGCCTGGTCTACGCCCGGAACCTGGCTGCTGTCCGACCTGCTGCTGACAGTCTCTGCCCC 180

Qy 181 CTGAAGGCGGTGAGGCGGTGAGGCTCCGGGGCCTGCGCCTCCTGCGGGGGCTTCCTGGGGCCGCCCTG 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 CTGAAGGCGGTGAGGCGGTGAGGCTCCGGGGCCTGCGCCTCCTGCGGGGGCTTCCTGGGGCCGCCCTG 240

Qy 241 GTCTTCGGGGTGGCCCACTTCTTCCCACTCCTATGCGGGGGCTTCCTGGGCCGCCCTG 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GTCTTCGGGGTGGCCCACTTCTTCCCACTCCTATGCGGGGGCTTCCTGGGCCGCCCTG 300

Qy 301 AGTGCAGGCGCTACCTGGGAGCAGCAGCCTTCCCTGGGCTACCAGCCTTCGGAGGCCG 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 AGTGCAGGCGCTACCTGGGAGCAGCAGCCTTCCCTGGGCTACCAGCCTTCGGAGGCCG 360

Qy 361 TGCTATTCCTGGGGGGTGTGCGCGGCCATCTGGGCCTCGTGTCGTGTCACCTGGGTCTG 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 TGCTATTCCTGCGGGGGGTGTGCGCGGCCATCTGGGCCTCGTGTCGTGTCACCTGGGTCTG 420

Qy 421 GTCTTTGGGTTGGAGGCTCCAGGAGGCTGGCTGGACCAACAACACCTCCCTGGGCATC 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421 GTCTTTGGGTTGGAGGCTCCAGGAGGCTGGCTGGACCAACAACACCTCCCTGGGCATC 480

Qy 481 AACACACCGGTCAACGGCTCTCCGGTCTCCTGAGGCCTCGGGACCCGGCCTCTGCCGGC 540
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 481 AACACACCGGTCAACGGCTCTCCGGTCTCCTGAGGCCTCGGGACCCGGCCTCTGCCGGC 540

Qy 541 CCGGCCCGCTTCAGCCTTCTCTCTGCTCTTTTTCTGCCCCTTGGCCATCATCACAGCCTTC 600
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 541 CCGGCCCGCTTCAGCGTCTTCTCTCTGCTCTTTTTCTGCCCCTTGGCCATCATCACAGCCTTC 600

Qy 601 TGCTACGTGGGCTGCCTCCGGGCACTGGCCGCGTCCGGCCTCGACGCACAGGCGGAAGCTG 660
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 601 TGCTACGTGGGCTGCCTCCGGGCACTGGCCGCGTCCGGCCTCGACGCACAGGCGGAAGCTG 660

Qy 661 CGGGCGCCTGGGTGGCCCGGCGGGGGCCCTCCTCACGCTGCTCTGCCTAGGACCCTAC 720
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 661 CGGGCCAAATGGGTGGCCCGGCGGGGGCCCTCCTCACGCTGCTCTGCCTAGGACCCTAC 720

Qy 721 AACGCCTCCAACGTGGCCAGCTTCCTGCTACCCCAATCTAGGAGAGGCTCCTGGCCGGAAGCTG 780
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 721 AACGCCTCCAACGTGGCCAGCTTCCTGCTACCCCAATCTAGGAGAGGCTCCTGGCCGGAAGCTG 780

Qy 781 GGGCTCATCACGGGGTGCCTGGAGTGTGGTGCTTAATCCGTGGTGACCGGTTACTTGGGA 840
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 781 GGGCTCATCACGGGGTGCCTGGAGTGTGGTGCTTAATCCGTGGTGACCGGTTACTTGGGA 840

Qy 841 AGGGGTCCTGGCCTGAAGACAGCAGTGTGCGGCCAAGAACGCAAGGGGCAAGTCCCAGAAG 900
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 841 AGGGGTCCTGGCCTGAAGACAGCAGTGTGCGGCCAAGAACGCAAGGGGCAAGTCCCAGAAG 900

Qy 901 TAA 903

Db 901 TAA 903

RESULT 7
ADC22790
ID   ADC22790 standard; cDNA; 903 BP.
XX
AC   ADC22790;
XX
DT   18-DEC-2003 (first entry)
XX
DE   Human G protein-coupled receptor cDNA #76.
XX
KW   Human; gene; ss; G protein-coupled receptor; GPCR;
KW   transmembrane-6 region; TM6; intracellular-3 region; IC3.
XX
OS   Homo sapiens.
XX
PN   US6655339-B1.
XX
PD   29-APR-2003.
XX
PF   13-OCT-1998; 98US-00170496.
XX
PR   14-APR-1997; 97US-00839449.
PR   14-APR-1998; 98US-00060188.
PR   26-JUN-1998; 98US-0090783P.
PR   07-AUG-1998; 98US-0095677P.
XX
PA   (AREN-) ARENA PHARM INC.
XX
PI   Liaw CW, Behan DP, Chalmers DT;
XX
DR   WPI; 2003-742861/70.

DR   P-PSDB; ADB61416.
XX
PT   Fatty acid and eicosanoid-binding G-protein coupled receptor protein
PT   GPR40 for control of pancreatic function and treatment of diabetes.
XX
PS   Claim 12; Page 241-242; 257pp; Japanese.
XX
CC   The invention relates to a novel screening method comprising a fatty acid
CC   and eicosanoid-binding G-protein coupled receptor protein, GPR40,
CC   originating in mouse, rat, crab-eating monkey and hamster (sequences
CC   fully defined in the specification), equivalent proteins of similar
CC   activity, and peptides containing partial sequences of the GPR40 protein.
CC   The novel fatty acid and GPR40 protein and their compositions have the
CC   following activities: antidiabetic, anabolic, neuroprotective,
CC   nephrotropic, ophthalmological, antilipaemic, dermatological,
CC   antiarteriosclerotic, antiarthritic, osteopathic, thrombolytic,
CC   nootropic, anorectic, hypotensive, and cytostatic. The novel fatty acid
CC   and eicosanoid-binding G-protein coupled receptor protein, GPR40, and
CC   further compositions can be used in the treatment, prevention and
CC   diagnosis of diabetes, ketosis, acidosis, diabetic neuropathy, diabetic
CC   nephropathy, diabetic retinopathy, hyperlipaemia, skin diseases,
CC   arthritis, bone diseases, arteriosclerosis, thrombosis, hypertension,
CC   indigestion, memory loss, obesity, hypoglycaemia, edema, insulin
CC   resistance, insulin allergy, fat metabolism disorders and cancer. This
CC   polynucleotide sequence represents the DNA encoding a monkey GPR40
CC   protein of the invention.
XX
SQ   Sequence 900 BP; 108 A; 322 C; 285 G; 185 T; 0 U; 0 Other;

Query Match   93.1%; Score 840.8; DB 9; Length 900;
Best Local Similarity   95.9%; Pred. No. 2.9e-155;
Matches 863; Conservative   0; Mismatches 37; Indels   0; Gaps   0;

Qy   1 ATGGACCTGCCCCGCAGCTCTCCTTCGGCCTGCTCTCATGTGCGCCGCCTTTGCGCTGGGCTTC 60
            ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   1 ATGGACCTGCCCCGCAGCTCTCCTTCGCCCTCATGTGCGCCGCCTTTGCGCTGGGCTTC 60

Qy   61 CCGCTCAAGTCTCTGGCCATCCGAGGCGCGACGGCCCAGCCCCGGCTCTCACCCCT 120
            ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   61 CCGCTCAAGTCTCTGGCCATCCGAGGCGCGACGGCGCCCAGCCCCGGCTCTCACCCCC 120

Qy   121 AGCCTGGTCTACGCCCTGACCTGGCTCGCTGCTCCGGGCCTGCTGCTGACAGTCTCTGCC 180
            ||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db   121 AGCCTGGTCTACGCCCTGACCTGGCTCGCTGCTCCGACCTGGTCTGACAGTCTCCTGCCC 180

Qy   181 CTGAAGGCGGTGGAGGCGGTGGCTAGCCTGGGCCTGCTGCGGCCCTGCGGCTGTGCCCC 240
            ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   181 CTGAAGGCGGTGGAGGCGGTGGCGGCGGCTGGGCCTGCTGCGGCCCTGCGGCTGTGCCCT 240

Qy   241 GTCTTCGCGTGGCCCACTTCTTCCCCACTCTATGCCCGGGGGGGCTTCCTGGCCGCCCTG 300
            ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   241 GTCTTCGGGGTGGGGTGGCCCACTTGCTCCACTCTATGCCCGGGGGGGGCTTCCTGGCCGCCCTG 300

Qy   301 AGTGCCAGGCCGCTACCTGGGAGCAGCAGCCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCG 360
            ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   301 AGTGCAGGCCGCTACCTGGGAGCGCTGCAGCTGGCTGGCTGGACCTTCCGGAGGCCG 360

Qy   361 TGCTATTCCTGGGGGTGTGCGCGACCTGGTCTG 420
            ||||||||||||||||||| ||||||||||||||
Db   361 TGCTATTCCTGGGGGTGTGTGCGCCATCACTGGGTCTG 420

Qy   421 GTCTTTGGGTTGGAGGGCTGCGGAGGCTGGACCACACCTCCCTGGGCATC 480
            ||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   421 GTCTTTGTGTTGGAGGCTGCCTGGACGCTGGACCACAACACCTCACTGGGCATC 480

Qy   481 AACACACCGGTCAACGGCTCAACGGCTCTCCCCGGCCTCTGCCGGC 540
            |||||||||||||||||||||| |||||||||||||||||||||||||
Db   481 AACACACCGGTCAACGGCTCAACGGCTCTCGGGGACCCGGCCTCTGCCGGC 540

Qy   541 CCGGCCCGCTTCAGCCTCTTCTCTTTTTCTGCCTTGGCCATCACAGCCTTC 600
            ||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   541 CCGGCCCGCTTCAGCCTCTTCTCTTCTTTTTCTGCCTTGGCCATCACAGCCTTC 600

Db   541 CCGGCCCGCTTCAGCCTCTCTCGCTGCTTCCGGGCACTGGCCGCCTGCTGGGCCGACTGGCCGCTCCGGGACTGGCCGCAGGCCGCGAAGCTG 600

Qy   601 TGCTACGTGGGCTGCCTCGGGGCACTGGCCCGGCTCCGGGCCTGACGACGCAGGCGGAAGCTG 660
            |||||| ||||||||||||||||| |||||||| |||||||||||||||||||||||||||||
Db   601 TGCTACGTGGGCTGCCTCGGGGCACTGGGCCGGGCTCCGGGCCTGACGCTGCCTAGGACCCTAC 660

Qy   661 CGGGCCGCTGGGGTGGCCGCGGGGGCCCTCCTCACGCGTGCTCTGCCGTCTGCTAGGACCCTAC 720
            |||||| ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   661 CGGGCCCAAATGGGTGGCCGCGGGGGCCCTCCTCCTCACGCGTGCTCTGCCGTCTGCTAGGACCCTAC 720

Qy   721 AACGCCTCCAACTGGCCAGCTTCCTGTACCCCAATCTAGGAGGCTCCTGGCCGGAAGCTG 780
            |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db   721 AACGCCTCCAACTGGCCAGCTTCCTGTACCCCAATCTAGGAGGCTCCTGGCCGGAAGCTG 780

Qy   781 GGGCTCATCACGGGCGTGCTGGAGTGTGGTGCTTAATCCGTGGTGACCGGTTACTTGGGA 840
            |||||| |||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   781 GGGCTCATCACGGGCGTGCTGGAGTGTGGTGCTTAATCCGTGGTGACCGGTTACTTGGGA 840

Qy   841 AGGGGTCCTGCCTGAAGAACAGTTGTGCGGCAAGAACCCAAGGGGCAAGTCCCAGAAG 900
            |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db   841 AGGGGTCCTGGCCTGAAGAACAGTTGTGCGGCAAGAACCCAAGGGGCAAGTCCCAGAAG 900

Qy   901 TAA 903
            ||
Db   901 TAA 903

RESULT 8
ADB61417
ID   ADB61417 standard; DNA; 900 BP.
XX
AC   ADB61417;
XX
DT   04-DEC-2003 (first entry)
XX
DE   DNA encoding monkey GPR40 protein.
XX
KW   fatty acid; eicosanoid-binding G-protein coupled receptor; GPR40;
KW   antidiabetic; anabolic; neuroprotective; nephrotropic; ophthalmological;
KW   antilipaemic; dermatological; antiarteriosclerotic; antiarthritic;
KW   osteopathic; thrombolytic; nootropic; anorectic; hypotensive; cytostatic;
KW   diabetes; ketosis; acidosis; diabetic neuropathy; diabetic nephropathy;
KW   diabetic retinopathy; hyperlipaemia; skin disease; arthritis;
KW   bone disease; arteriosclerosis; thrombosis; hypertension; indigestion;
KW   memory loss; obesity; hypoglycaemia; edema; insulin resistance;
KW   insulin allergy; fat metabolism disorder; cancer; monkey; gene; ds.
XX
OS   Primates.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..900
FT   /*tag= a
FT   /partial
FT   /note= "No stop codon"
FT   /product= "Monkey GPR40 protein"
XX
PN   WO200306895-A1.
XX
PD   21-AUG-2003.
XX
PF   13-FEB-2003; 2003WO-JP001483.
XX
PR   14-FEB-2002; 2002JP-00037131.
PR   12-JUL-2002; 2002JP-00204163.
PR   12-NOV-2002; 2002JP-00328696.
PR   22-JAN-2003; 2003JP-00014032.
XX
PA   (TAKE) TAKEDA CHEM IND LTD.
XX
PI   Hinuma S, Hosoya M, Ito Y, Kobayashi M, Tanaka H, Okubo S;
PI   Fujii R, Kizawa H, Kawamata Y, Ogi K, Harada M, Fukusumi S;
XX
DR   WPI; 2003-671661/63.

The invention relates to a novel screening method comprising a fatty acid and eicosanoid-binding G-protein coupled receptor protein, GPR40, originating in mouse, rat, crab-eating monkey and hamster (sequences fully defined in the specification), equivalent proteins of similar activity, and peptides containing partial sequences of the GPR40 protein. The novel fatty acid and GPR40 protein and their compositions have the following activities: antidiabetic, anabolic, neuroprotective, nephrotropic, ophthalmological, antilipaemic, dermatological antiarteriosclerotic, antiarthritic, osteopathic, thrombolytic, nootropic, anorectic, hypotensive, and cytostatic. The novel fatty acid and eicosanoid-binding G-protein coupled receptor protein, GPR40, and further compositions can be used in the treatment, prevention and diagnosis of diabetes, ketosis, acidosis, diabetic neuropathy, diabetic nephropathy, diabetic retinopathy, hyperlipaemia, skin diseases, arthritis, bone diseases, arteriosclerosis, thrombosis, hypertension, indigestion, memory loss, obesity, hypoglycaemia, edema, insulin resistance, insulin allergy, fat metabolism disorders and cancer. This polynucleotide sequence represents the DNA encoding a hamster GPR40 protein of the invention.

```
XX
CC   The invention relates to a novel screening method comprising a fatty acid
CC   and eicosanoid-binding G-protein coupled receptor protein, GPR40,
CC   originating in mouse, rat, crab-eating monkey and hamster (sequences
CC   fully defined in the specification), equivalent proteins of similar
CC   activity, and peptides containing partial sequences of the GPR40 protein.
CC   The novel fatty acid and GPR40 protein and their compositions have the
CC   following activities: antidiabetic, anabolic, neuroprotective,
CC   nephrotropic, ophthalmological, antilipaemic, dermatological,
CC   antiarteriosclerotic, antiarthritic, osteopathic, thrombolytic,
CC   nootropic, anorectic, hypotensive, and cytostatic. The novel fatty acid
CC   and eicosanoid-binding G-protein coupled receptor protein, GPR40, and
CC   further compositions can be used in the treatment, prevention and
CC   diagnosis of diabetes, ketosis, acidosis, diabetic neuropathy, diabetic
CC   nephropathy, diabetic retinopathy, hyperlipaemia, skin diseases,
CC   arthritis, bone diseases, arteriosclerosis, thrombosis, hypertension,
CC   indigestion, memory loss, obesity, hypoglycaemia, edema, insulin
CC   resistance, insulin allergy, fat metabolism disorders and cancer. This
CC   polynucleotide sequence represents the DNA encoding a hamster GPR40
CC   protein of the invention.
XX
SQ   Sequence 900 BP; 122 A; 312 C; 267 G; 199 T; 0 U; 0 Other;

Query Match             67.8%;  Score 612;  DB 9;  Length 900;
Best Local Similarity   80.0%;  Pred. No. 1.5e-110;
Matches  720;  Conservative    0;  Mismatches  180;  Indels    0;  Gaps    0;

Qy         1 ATGGACCTGCCCCCGCCAGCTCTCCTTCGGCCTCGGCCCTCATGTGGCGCCCTTGCGCTGGGGCTTC 60
             ||||| ||| ||| ||| |||||| |||| |||||||||||||| ||||||| |||||||||||||||
Db         1 ATGGGCCCGTCTCCCCAACTCTTCTTCTGCCCTCATGTGTCTGCCCTCGGCGCTGGGGCTTC 60

Qy        61 CCGGTCAACGTCCTGCCATCGAGGCGCGACGGCCCACGCCCGGCTCCGGTCCGTCACCCCT 120
             |||| | |||| |||||||||||||| || ||| ||| |||| ||||||||| ||||||||
Db        61 CCGGCTGAACCTGTTGCCATCGAGGCGCCGTGCCGGGCGCGGGCGGCTCCGGTCACCCCCC 120

Qy       121 AGCCTGGTCACGCCCGAACCTGGCCGCTCCTGCGCTGCGTGGCCGCTGACCAGTCTCTGCCC 180
             | |||||||| |  |||||||||||||| |||||| | ||||| |||| |||| |||| |||
Db       121 AACCTGGTCTATACACTCCACCTGGCCGTCTCGAACCTGGCTCTGCGGCCATCACGCTACCC 180

Qy       181 CTGGAGGCGGTGGAGGGCTTCCGGGGCGCTAGCCTCCGGGGCCCTGCGGGGCTGTGCCCC 240
             |||||||  ||| ||||||||| |||||| |||| |||| |||| |||  ||||| ||||
Db       181 GTGGAAGGCCGTGCCGGGCGCGGGCGGCTCCGGTCCTGGGTACCCGGTTCCCCGTTGTGCCCT 240

Qy       241 GTCTTCGCGGTGGCCCACTTCTTCCACTCTATGCGCCGGGGGCTTCCTGGCCGCCGCTG 300
             |||| | || ||| ||||| ||| || |||||||||| ||| ||||||||||||| ||
Db       241 GTCTTTGTCTTGGTGCACTCCCCCATGCGGCGGGGGGCTTCCTGGCCGGAGGCCTCTC 300

Qy       301 AGTGCGGCCTACGTGGTGCGGGAGCAGCCTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCG 360
             |||| | |||| || |||||||||||||| |||||||||| |||||||| ||||||||||||
Db       301 AGTGCTGGCCTACCTGGGTGCTTCGGGGGGGCTGCCTTCCCGGTTCCCCCGTTCGGGCGCGCCC 360

Qy       361 TGCTATTCCTGGGGGGTGTGCGGCCATCGGACCTCTGTCCTGTCACCTGGGCTG 420
             ||| || | |  ||| || || || ||| || ||||| | ||| || ||||| ||
Db       361 CGCTACTCCTGGGGCGTGTGTGGGCGTATATGGGGGTGTGCCGTCCACATGGGCGTG 420

Qy       421 GTCTTTGGGTTGGAGCTGCTCAGGAGGCTGGCTGGACCACAGCAACACCTCCCTGGCATC 480
             |||| ||| |||||||| ||||| ||| ||| | ||||| || ||||| |||||||||  |
Db       421 GTCTTTGGGCTGGAGAGTGCGTCCGGGAGCGTGCTGGACGCAGCTGCGGGTCCCTGGGAATC 480

Qy       481 AACACCGGTCAACGCCATGCGGCTCTGGGAGGCCTGGGACGGTCGGGCTCCGGGC 540
             ||||| ||| ||||  |||| |||||||||||||| |||||||  ||||||||||
Db       481 AACACACCGGTGGAATGGTTCCCCGGTGTGCGGGAGGGCCTGGGGATCCCAACTCTGCCGCGG 540

Qy       601 TGCTACGTGGCTGGCTGCCTCCGGGCACTGGCCCGCGCTCCGGCCTGACGCAGGCGGAAGCTG 660
             ||| ||||||| |||||||||||||||||||||  |  | |||||||||||||||||||||||
Db       601 TGCTACGTGGGCTGGCTGCCTCCGGGCACTGGCCGCCCTGACGCAGGCGGAAGCTG 660

Qy       661 CCGGCCGGCCTGGGTGGCCGGGGGGCCCTCCTCACGCGCTGCTGTCTGCGGTAGGACCCTAC 720
             | |||||||||||||| ||||||||||| ||||||||||||||||  | ||||||||||||||
Db       661 AGGGCCGCCCTGGGTAGCCGGGGGGCCGGGGGGCCCTCCTCACGCGCTGCTGCGGTAGGACCCTAC 720

Qy       721 AACGCCTCCAACGTGGCCAGCTTCTGTACCCCAATCTAGGAGGCTCCTGGCGGAAGCTG 780
             ||| | ||||| ||||||||||||||  |||||||| ||||||||||||||  ||||||
Db       721 AACCTGGTCTCCAATGGCCAGCTTCTGAACCCCAATCTCGGAGGCTCCTGGCCGGAAGCTG 780

Qy       781 GGGCTCATCACGGGTGCCTGGAGTGTGGTGCTTAATCCGCTGGTGACCGGTTACTTGGA 840
             |||||||||||||||||||||||||||||||| |  |||||||||||||||||||| ||
Db       781 GGGCTCATCACGGGTGCCTGGAGTGTGGTGCTCACCCCGGTGGTGACCGGTTACTTGGGA 840

Qy       841 AGGGGTCCTGGCCTGAAGAGACAGTGTGTGGCGGCAAGAACCCAAGGGGCAAGTCCCAGAAG 900
             |||||||||||||||||||||  | ||||| |  |||||||| |||||||| |||||||||
Db       841 AGGGGTCCTGGCCTGAAGAGAACCAAGGAGCACGTCCCAGAAG 900
```

```
RESULT 9
ADB61429
ID   ADB61429 standard; DNA; 900 BP.
XX
AC   ADB61429;
XX
DT   04-DEC-2003 (first entry)
XX
DE   DNA encoding hamster GPR40 protein.
XX
KW   fatty acid; eicosanoid-binding G-protein coupled receptor; GPR40;
KW   antidiabetic; anabolic; neuroprotective; nephrotropic; ophthalmological;
KW   antilipaemic; dermatological; antiarteriosclerotic; antiarthritic;
KW   osteopathic; thrombolytic; nootropic; anorectic; hypotensive; cytostatic;
KW   diabetes; ketosis; acidosis; diabetic neuropathy; diabetic nephropathy;
KW   diabetic retinopathy; hyperlipaemia; skin disease; arthritis;
KW   bone disease; arteriosclerosis; thrombosis; hypertension; indigestion;
KW   memory loss; obesity; hypoglycaemia; edema; insulin resistance;
KW   insulin allergy; fat metabolism disorder; cancer; hamster; gene; ds.
XX
OS   Cricetinae gen. sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..900
FT                   /tag= a
FT                   /partial
FT                   /note="No stop codon"
FT                   /product= "Hamster GPR40 protein"
XX
PN   WO200306B959-A1.
XX
PD   21-AUG-2003.
XX
PF   13-FEB-2003; 2003WO-JP001483.
XX
PR   14-FEB-2002; 2002JP-00037131.
PR   12-JUL-2002; 2002JP-00204163.
PR   12-NOV-2002; 2002JP-00328696.
PR   22-JAN-2003; 2003JP-00014032.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Hinuma S, Hosoya M, Ito Y, Kobayashi M, Tanaka H, Okubo S;
PI   Fujii R, Kizawa H, Kawamata Y, Ogi K, Harada M, Fukusumi S;
XX
DR   WPI; 2003-671661/63.
DR   P-PSDB; ADB61428.
XX
PT   Fatty acid and eicosanoid-binding G-protein coupled receptor protein
PT   GPR40 for control of pancreatic function and treatment of diabetes.
XX
PS   Claim 13; Page 246-247; 257pp; Japanese.
```

```
CC   reporter gene expression; or (b) detecting whether the compound binds to
CC   GPR40 and increases glucose-stimulated insulin release from mammalian
CC   pancreatic beta cells. The method is useful for identifying GPR40
CC   antagonist or agonist compounds for treating disorders e.g. type 2
CC   diabetes, obesity, glucose or insulin intolerance, neurodegenerative
CC   disease (e.g. Alzheimer's disease) or stroke. The present sequence is the
XX   mouse cDNA for GPR40.
SQ   Sequence 903 BP; 138 A; 300 C; 245 G; 220 T; 0 U; 0 Other;

Query Match             66.9%;  Score 604.4;  DB 6;  Length 903;
Best Local Similarity   79.4%;
Matches  716;  Conservative    0;  Mismatches  186;  Indels    0;  Gaps    0;

Qy         1 ATGGACCTGCCCCCAGCCTCCTTCGGCCTCTCATGTGGCCGCCTTGCCTGCTGGGCTTC 60
             |||||||||||||| || || |||| ||||||||| | ||  ||||||||||||| ||||
Db         1 ATGGACCTGCCCCCACCAGCTCTCCTTCGCTCTCATGTATCTGCCTTTGCCTGGCTTT 60

Qy        61 CCGGCTCAACGTCCTGCCATCCGAGGCGCGACGGCCCACGCCGCCTCCTCACCCCT 120
             ||  |||||| || | |||||||||||| | | | ||||||| |||  ||| || |||
Db        61 CCATTGAACTTGTTAGCCATCCGAGGCGCAGCTGTCCCACGCTAAACTGCGACTCACTCCC 120

Qy       121 AGCCTGGTCTACGCCCTGAACCTGGGCTGCTCCGACCTGCTCTGCGACAGTCTCTCTGCCC 180
             ||| ||||||||| |||  ||| ||||| |||| | | |||| ||| ||| | || |||||
Db       121 AGCTTGGTCTACACTCTCCATCTGGGCGCTGCTCTGATCTCCTACTGCTCATCACTCTGCGCC 180

Qy       181 CTGAAGGCGGTGGAGGCGTAGCCTCCGGGCTCCGGGCTTGGCCTCGCCGCGTGTGCCCC 240
             ||||||||| |||||||| |  || |  |  |||  | |||||| || ||| | ||||
Db       181 CTGAAGGCGTGGAGGCGCCTGCTGTGTCTGAGCTGCTGGGCCCTGCGCCCTTCTGCCCA 240

Qy       241 GTCTTCGCGGTGGCCCACTTCTTCCCACTCTATGCCGGGGGGCTTCCTGGCCGCCCCCTG 300
             ||||| ||| || ||||| || | |||| |  || | | | | ||||||| || || ||
Db       241 GTCTTTGCCTTGCCCACTTTGCTCCCGGTCTCCCCTTCTCTCTAGCGTCGTCTC 300

Qy       301 AGTGCAGGCGCTACCTGGGAGCAGCCTTCCCCTTGGGTTACCAAGCCTTCCGGAGGCCG 360
             ||  | || ||||||||| |||| |  | | ||| ||||||| |||||  |||||| ||
Db       301 AGCGGCTGGCCGCTACCTGGGGGCGCCGGGCTGCGGTTACCCAGTACCCATCCGGAGCGCCC 360

Qy       361 TGCTATTCCTGGGGGTGTCGCGGCCATCCTGGGCCCTCGTCCTCCTGCACCTGGGTG 420
             ||||||||||||||||||   | | || | |||||| |||||| ||||||||||| ||
Db       361 CGCTATTCCTGGGGGTGTGTGTGTGCTGCTGTATATGGGCCTTGTCCTCTGCCACCTGGGGCTG 420

Qy       421 GTCTTTGGGTTGGAGGCTCCAGGAGCTGGCTGCTGCTCTGGAGGCCTGGGCCATC 480
             | ||| ||| ||||| | |  ||| ||| | | || || || || || |||||||
Db       421 GCCCTTGGCTTGGAGGACTTCCGGAAGCTGCGCTGGCTGAACAAGTACCAGTTCCTGGGCCATC 480

Qy       481 AACACACCGGTCAACGGCTCTCCGGTCTGGGCTGCTGGAGCCTGGCCTCGCCGGC 540
             ||| | |||| || |||| || |||| | || ||||||||||||  | |  || |
Db       481 AACATACCCGTGAATGGCTCCCCGGTGCTGCGCGGCTGGGAGCCTGGGATCCCGACTCTCGCCCGC 540

Qy       541 CCGGCCCGCTTCAGGCTCTCAGCTCTCTTTTTCTGCCCTGGCCATCACAGCCTTC 600
             || || ||| | |   | |||| | |||  ||| |||||||| |||||| ||||||
Db       541 CCTGCCCGTCTCAGTTTCTCCATTCTGCCTTCTTTCTGCCCTGCTCATCACTGCCTTC 600

Qy       601 TGCTACGTGGGCTGCTCCTCCGGCCACTGGGCCCTGACGCACAGGCGGAAGCTG 660
             ||||| ||||||||||||||||||| || ||||||| ||| ||| |||||||| 
Db       601 TGCTATGTGGGCTGCTCCTCCGGCCCTGCCCTGGCGCCTGAGCCACAAACGGAAGCTC 660

Qy       661 CGGGCCGCCTGGGTGGCCGGCGGGGCCCTCCAGCTGCTCCGTGCTGCTGCGTAGGACCCTAC 720
             ||||| || ||||| |||||||| | || ||||| || ||||| ||||| || |||| ||||||  
Db       661 AGGGCAGCCTGGGTGGCCGGAGGGCCGCTCTCCGTGCTGCTGCGTCTGCGTGGGGCCCTAT 720

Qy       721 AACGCCTCCAACGTCGCCCAGCTTCCTGTACCCCAATCTAGGAGGCCTCGGCCGGAAGCTG 780
             || ||||||||  | || |||| ||   ||  |||||| || ||| || |||| |||||
Db       721 AATGCCTCCAATGTGGCTAGTTTCATAAACCCGACCTAGGAGGGCTCCTGGAGGAAGTTG 780

Qy       781 GGGCTCATCACGGGTGCTCGGAGTGTGGTGTCGGTTAATCCGCTGGTGACCGGTTACTTGGGA 840
             || ||||||||  ||| | | ||| ||| | || | |||| |||||| |||  | |||||||
Db       781 GGACTCATCACAGGGCCTTGGGAGTTGGTGACTCAGTCCACCCCACTGGTCACTCGGTACTTGGGA 840

Qy       841 AGGGGTCCTGCCTGAAGACAGTGTGCGGCAAGAACGCAAGGGGCAAGTCCCAGAAG 900
             | ||| ||| | |||||| || | |  | |||||| | ||||| |||   | |||||
Db       841 ACAGGTCCTCGACGCGGGAACAATATGTGTGACGAGGAACAATTCAGAAG 900
```

```
Db       661 AGGGCAGCCTGGGCGCGGGCCGAGGGGCCTTTCTACACTCCGCTGCTGGGGCCCTAC 720
Qy       721 AACGCCCCAACTGGCCAGCTTCCTGTACCCCAATCTAGGAGGCCTGGCCGAAGCTG 780
Db       721 AATGCCTCCAATGTGGCAGTTTCGTAAACCCGACCTGGGAGGCTCCTGGACGAAGCTG 780
Qy       781 GGGCTCATCACGGGTGCCTGGAGTGTGGCTTAATCCGCTGGTGACCGGTTACTTGGGA 840
Db       781 GGGCTCATCACAGGGTCCTGAGTGTGGTGGCTCCAACCCGCTGGTCACCGGTCACTTGGGA 840
Qy       841 AGGGGTCCTGGCCTGAAGACAGTGTGCGGCAAGAACGCAAGGGGCAAGTCCCAGAAG 900
Db       841 GCAAGTCCTGGCCCAGGGACCAGCAGTATGTACGACAAGGACTCAAGGAGGAACAATTCAGAAG 900

RESULT 10
ABK90237
ID   ABK90237 standard; cDNA; 903 BP.
XX
AC   ABK90237;
XX
DT   21-OCT-2002 (first entry)
XX
DE   Mouse cDNA encoding G protein-coupled receptor GPR40.
XX
KW   Mouse; ss; gene; GPR40; G protein-coupled receptor; type 2 diabetes;
KW   obesity; antidiabetic; neuroprotective; anorectic; cerebroprotective; Gq;
KW   G protein; reporter gene; glucose intolerance; insulin intolerance;
KW   neurodegenerative disease; Alzheimer's disease; stroke.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..903
FT                   /*tag= a
FT                   /product= "GPR40"
XX
PN   WO200257783-A2.
XX
PD   25-JUL-2002.
XX
PF   18-DEC-2001; 2001WO-US048985.
XX
PR   22-DEC-2000; 2000GB-00031527.
XX
PA   (GLAX ) GLAXO GROUP LTD.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI   Andrews JL, Briscoe CP, Ignar DM, Muir AI, Sauls HR, Tadayyon M;
XX
DR   WPI; 2002-599726/64.
DR   P-PSDB; ABG31107.
XX
PT   Identifying GPR40 receptor ligand for treating disorders e.g. obesity,
PT   comprises detecting whether the test compound competitively inhibits the
PT   binding of a fatty acid GPR40 ligand to a GPR40 receptor.
XX
PS   Disclosure; Page 49-50; 53pp; English.
XX
CC   The invention relates to screening a test compound to determine whether
CC   the compound is a GPR40 receptor (G protein-coupled receptor) ligand
CC   comprises detecting whether the test compound competitively inhibits the
CC   binding of a fatty acid GPR40 ligand to a GPR40 receptor. Also included
CC   are (1) a method of screening a compound for GPR40 antagonist activity,
CC   comprising: (a) measuring any detectable signal produced by a reporter
CC   gene (comprising a reporter gene under the control of G protein Gq
CC   responsive transcriptional element); (b) detecting a decrease in reporter
CC   gene expression in the presence of both test compound and agonist; or (c)
CC   detecting whether the compound decreases glucose-stimulated insulin
CC   release from mammalian pancreatic beta cells in the presence of a GPR40
CC   agonist, compared to glucose-stimulated insulin release that would occur
CC   due to the presence of the GPR40 agonist; (2) a method of screening a
CC   compound for GPR40 agonist activity, which comprises: (a) detecting any
```

```
CC   nephropathy, diabetic retinopathy, hyperlipaemia, skin diseases,
CC   arthritis, bone diseases, arteriosclerosis, thrombosis, hypertension,
CC   indigestion, memory loss, obesity, hypoglycaemia, edema, insulin
CC   resistance, insulin allergy, fat metabolism disorders and cancer. This
CC   polynucleotide sequence represents the DNA encoding a mouse GPR40 protein
CC   of the invention.
XX
SQ   Sequence 900 BP; 137 A; 300 C; 244 G; 219 T; 0 U; 0 Other;

Query Match             66.7%;  Score 602.4; DB 9;  Length 900;
Best Local Similarity   79.3%;  Pred. No. 1.1e-108;
Matches  714;  Conservative    0;  Mismatches 186;  Indels    0;  Gaps    0;

Qy            1 ATGGACCTGCCCCCGGAGCTCCTTCGGCCTCCTCTATGTGGCGCCCCTTGCCGCTGGGCTC 60
                ||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db            1 ATGGACCTGCCCCCAGCTCCTTCCTTCGCTCTTCCTATGTATCTGCCTTGCCGCTGGGCTTT 60

Qy           61 CCGCTCAACGTCCTGGCCATCGGAGGGCGACGGCCCACGCCCGGCTCCGTCCACCCCTCACCCCT 120
                ||||| ||||||||||||||||||||||| ||||||||| |  ||||||||||||||||||
Db           61 CCATTGAACTTGTTAGCCATCCGAGGCGCAGTGTCCCACGTAAACTGCGACTCACTCCC 120

Qy          121 AGCCTGGTCACGCCCTGAACCTGGGCTGCTCCGACCTGCTGACAGTCTCTCGCCC 180
                || ||||||||| || ||  ||||||| |||||| ||  ||| |  ||||||||||
Db          121 AGCTTGGTCTCACACTCTCCATCGGGCGTCTGATCTCCACTGGCCATCACTCTGCCC 180

Qy          181 CTGAAGGCGGTGGAGGCGTAGCCTCCGGGCCTGGCCTCTGCGGCCTCGTGTGCCCC 240
                ||||||||||||||||||||| |||||||| |||| ||||||||| ||||||||||
Db          181 CTGAAGGCTGTGGAGGCCCTGCCTTCTGGAGCCTGGCCTGGCCCCGCTCCCCTTCTGCCCA 240

Qy          241 GTCCTCGCGGTGGCCCCACTTCTTCCCACTCTCATGCCGCCGGGGGCTTCCTGGCCGCCTG 300
                |||| |||||| |||   ||||| | ||||| |  ||||||||||| ||||||| ||| ||
Db          241 GTCTTTGCCTTGGCCCCACTTTGCTCCCCCTCACGCAGGGCGGAGGCTTCCTAGCTGTGCTCTC 300

Qy          301 AGTGCAGGCGCCGCTACCTGGGAGCAGCTCCCGGGCTGCGGCCGTCCTGTGTCCGTCTGTCACCTGGGTCTG 360
                |  | |||||||||||||||||||||| ||||| ||||||||| || |||||||||||||
Db          301 ACGGCTGGCGCCGCTACCTGGGGCCTGCTTCCCGGGTACCGAGGCCC 360

Qy          361 TGCTATTCCTGGGGGTGTGCGCGGCCATCTGGGCGTCCTGTGTCACCTGGGTCTG 420
                || ||||||||||||||||||||||||| ||||| ||||| | |||||||||||
Db          361 CGCTATTCCTGGGGTGTGTGTGGCTATATGGGCCCTTGTCTCTGTCTCTGCCACTGGGGGCTG 420

Qy          421 GTCTTTGGGTTGGAGGTCCAGGAGGCTGGCACCGGCAACACCTCCCTGGCATC 480
                ||||||||| |||||| |||||| |||  ||  |||||| |  ||||||||||
Db          421 GCCCTTGGACTTCCGGAACTGGCTGGACAACAGTACCAGTTCCCTGGGCATC 480

Qy          481 AACACACCGGTCAACGGCTTCACGGCTTCCGGTCTCCGGTCCGGACCCGGCCTCTGCCGGC 540
                |||||||||| |||||||| |||||||| ||||||||||||||||||| |||||||||||
Db          481 AACATACCCGTGAATGCTCCCGGCTCTCCCGGTCCGCGGATCCGGCTGGATCCGGCCGCGC 540

Qy          541 CCGGCCCGCTTCAGCCTCTCTCTCCTTTTTCTGCCTCGGCCATCACAGCCTTC 600
                |||||| |||||||||| || ||||||| ||||||| |||||||||||||||||
Db          541 CCTGCCCGTCTCAGTTTCTCCATTCTTCTCTTCTTCTGCGCTCTGGTCATCGCCTTC 600

Qy          601 TGCTACGTGGGCTGCCTCCGGGCACTGGCCCCTGGTGCGCTGCCCTCAGGCGGAAGCTC 660
                ||||||||| |||||||||||||||||||||||||||  || ||||||||| |||||||
Db          601 TGCTATGTGGGCTGCCTCCGGGCCCGGGGGCGCCTGGTGCGCTGCCTCAGGCCTGAGCTC 660

Qy          661 CGGGGCCGCTGGGGTGGCCGGCAGCGGCCGCTCCTCACCGTGCTCTCGCGTAGGACCCTAC 720
                ||||||||||||||||||||||||||||||||||||||||||||| ||| |  ||||||||
Db          661 AGGGCAGCTTGGGTGGCCGGAGGCGGCCGCTCCTCCTCCACACTCCTGCCTCTGCCTGGGCCTCCCAT 720

Qy          721 AACGGCTCCAACGTGGCCAGCTTCCTGTACCCCAATCTAGGAGGCTCTGGCGGAAGCTG 780
                |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db          721 AATGCCCCAATGTGGCTAGTTCATAAACCCGGAACTTCCTGAGGAGGCTCCTGAGGAAGCTG 780

Qy          781 GGGCCTCATCACGGGTGCTGCCTGGAGTGTGGTGCTTAATCCGTCGGTTACTTGGA 840
                |||||||||||||||||||||||| |||||||||||| |  |||||| | |||||||
Db          781 GGACTTCATCACAGGGCCTGGAGTGTGGTACTCAACCCACTGGTCACTGGCTACTTGGA 840

Qy          841 AGGGGTCCTGCCCTGAAGACAGTGTGCGGCAAGAACGCAAGGGGCCAAGTCCCAGAAG 900
                || |||||||||||||| ||| |||||||||||||| || ||||| ||||||||||||
Db          841 ACAGTTCCTGACGCGGGAACAATATGTGTGACGAGGACTCAAAGAGGAACAATTCAGAAG 900
```

RESULT 12
ADB61403
ID   ADB61403 standard; DNA; 900 BP.
XX
AC   ADB61403;
XX
DT   04-DEC-2003 (first entry)
XX
DE   DNA encoding rat GPR40 protein.
XX
KW   fatty acid; eicosanoid-binding G-protein coupled receptor; GPR40;
KW   antidiabetic; anabolic; neuroprotective; nephrotropic; ophthalmological;
KW   antilipaemic; dermatological; antiarteriosclerotic; antiarthritic;
KW   osteopathic; thrombolytic; nootropic; anorectic; hypotensive; cytostatic;
KW   diabetes; ketosis; acidosis; diabetic neuropathy; diabetic nephropathy;
KW   diabetic retinopathy; hyperlipaemia; skin disease; arthritis;
KW   bone disease; arteriosclerosis; thrombosis; hypertension; indigestion;
KW   memory loss; obesity; hypoglycaemia; edema; insulin resistance;
KW   insulin allergy; fat metabolism disorder; cancer; rat; gene; ds.
XX
OS   Rattus sp.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..900
FT   /tag= a
FT   /partial
FT   /note= "No stop codon"
FT   /product= "Rat GPR40 protein"
XX
PN   WO2003068959-A1.
XX
PD   21-AUG-2003.
XX
PF   13-FEB-2003; 2003WO-JP001483.
XX
PR   14-FEB-2002; 2002JP-00037131.
PR   12-JUL-2002; 2002JP-00204163.
PR   12-NOV-2002; 2002JP-00328696.
PR   22-JAN-2003; 2003JP-0001 4032.
XX
PA   (TAKE) TAKEDA CHEM IND LTD.
XX
PI   Hinuma S, Hosoya M, Ito Y, Kobayashi M, Tanaka H, Okubo S;
PI   Fujii R, Kizawa H, Kawamata Y, Ogi K, Harada M, Fukusumi S;
XX
DR   WPI; 2003-671661/63.
DR   P-PSDB; ADB61402.
XX
PT   Fatty acid and eicosanoid-binding G-protein coupled receptor protein
PT   GPR40 for control of pancreatic function and treatment of diabetes.
XX
PS   Example 3; Page 234; 257pp; Japanese.
XX
CC   The invention relates to a novel screening method comprising a fatty acid
CC   and eicosanoid-binding G-protein coupled receptor protein, GPR40,
CC   originating in mouse, rat, crab-eating monkey and hamster (sequences
CC   fully defined in the description), equivalent proteins of similar
CC   activity, and peptides containing partial sequences of the GPR40 protein.
CC   The novel fatty acid and GPR40 protein and their compositions have the
CC   following activities: antidiabetic, anabolic, neuroprotective,
CC   nephrotropic, ophthalmological, antilipaemic, dermatological,
CC   antiarteriosclerotic, antiarthritic, osteopathic, thrombolytic,
CC   nootropic, anorectic, hypotensive, and cytostatic. The novel fatty acid
CC   and eicosanoid-binding G-protein coupled receptor protein, GPR40, and
CC   further compositions can be used in the treatment, prevention and
CC   diagnosis of diabetes, ketosis, acidosis, diabetic neuropathy, diabetic
CC   nephropathy, diabetic retinopathy, hyperlipaemia, skin diseases,
CC   arthritis, bone diseases, arteriosclerosis, thrombosis, hypertension,
CC   indigestion, memory loss, obesity, hypoglycaemia, edema, insulin
CC   resistance, insulin allergy, fat metabolism disorders and cancer. This
CC   polynucleotide sequence represents the DNA encoding a rat GPR40 protein
CC   of the invention.

XX
SQ   Sequence 900 BP; 142 A; 287 C; 244 G; 227 T; 0 U; 0 Other;

Query Match   65.1%; Score 588; DB 9; Length 900;
Best Local Similarity   78.3%; Pred. No. 7.3e-106;
Matches 705; Conservative   0; Mismatches 195; Indels   0; Gaps   0;

Qy   1 ATGGACCTGCCCCCGCGAGCTCTCCTTCGGCCTCTATGTGCCGGCTCGGCTTC 60
              |||||| | ||| ||||| ||||| |||||| |||||||||||||||||||
Db   1 ATGGACCTGCCCCCACAGCTCTCCTTCGCTCTCTATGTATCAGCCCTGGGCTTT 60

Qy   61 CCGCTCAACGTCCTGGCCATCCGAGGCGCACGGCCCACGCGCCTCCGTCCACCCCT 120
              ||| || | ||||| |||| ||||| || || ||||| |||||||| |||||
Db   61 CCATTGAACTTGTTAGCCATCCGAGGTGCAGTGTCCCACGGGAAACTGCCGACTCACCCCC 120

Qy   121 AGCCTGGTCACGCCCTGAACCTGGCTGCTCCGACCTGCTCTGACAGTCTCTGCCC 180
              ||| || || || |||||||||||||| | || |||||||| ||||||||||
Db   121 AGCTTGGTCTCACACTTCCATTGGCCGTCTCGACCTCCTACTGGCCATCACCCTGCCC 180

Qy   181 CTGAAGGCGGTGGAGGGCTAGCCTCCGGGCCTTCGGGCCCCGGCCCCTGCGCCCCTG 240
              |||||| ||||| |||||||||| ||||| ||||||||||||| |||| ||||||
Db   181 CTGAAGGCTGTGGAGGCCCTTGGCTTGGGGTCTGGGGTCTGCCCACTCCCCTTCTGCCCA 240

Qy   241 GTCTTCGCGGTGGCCCACTTCTTCCACACTTCCGGCGGGCCTTCTGGCCGCCCTG 300
              |||||||| | ||||||||||||| ||||||||| ||||||||||||||||||||
Db   241 GTCTTTGCCTTGGCCCACTTTGCCCCTCTATGCGGGTCCTTCTGGCTGCTCTC 300

Qy   301 AGTGCAGGGCCGCTACCTGGGAGCAGCTCCCTTTGGCCTTCCGGAGGCCG 360
              ||||||||||| ||||| |||||| ||||| | | |||| |||||||||||
Db   301 AGTGCTGGCGCTACCTGGGAGCTGCTCCCTTTGGATACCAAGCCATCCGGAGGCCC 360

Qy   361 TGCTATTCCTGGGGGTGTGCGGCCGGCCATCTGGGCCCTCGTCTGTGTCACCTGGGTCTG 420
              ||||| |||| ||| || ||| | | | | |||||||| || | | | || ||||| ||
Db   361 TGCTATTCCTGGGGTGTGTGTGCTGTATATGGGGCCCTTGCTTCCTTGCCCACCTGGGACTG 420

Qy   421 GTCTTTGGGTTGGAGGCTCCAGGAGGCTGGCTGGACCAGCAACACCTCCTGGCATC 480
              ||| |||| ||||||||| ||| |||||||| | ||||||| || ||||| |||
Db   421 GCTCTTCTGGGCTGGAGGCTCCAGAGGCTGGGTGATAAACACCACCAGTTCCCTGGGCATC 480

Qy   481 AACACCGGTCAACGGCTCTGGAGGCCTCCCCGTGGGACCGGCCTTCTCGCGGC 540
              ||||| ||| || |||| |||| |||||||||| |||||||| ||||| | |
Db   481 AACATACCCGTGAATGGCTCCCCCGTGGGTGAAGCGTGGATCCGCGGATCCTGACTTCGCCGC 540

Qy   541 CCGGCCCGCTTCAGCCTCTTCTTTTTCTGGCCATCACAGCCTTC 600
              ||||| || | |||| | |||| ||||| ||| ||| | ||||
Db   541 CCTGCCGACTCAGTTTCTGCCTCTTCTTCTTGTTTATCACTGCTTTC 600

Qy   601 TGCTACGTGGCTGCTCCGCACTGGGGCCCTGGTGCACTCGGGCCACAAACGAAGCTC 660
              |||| ||||||| | ||| ||| ||||| || | || ||| ||| ||||| ||||| |
Db   601 TGCTATGTGGCTGCTCCCGTGCACTCGGGCTCGAGCCACAAACGGAAGCTC 660

Qy   661 CGGGCCGCCTGGGTGGCCCGGGCGGGGTCCTGCTCTCGCGTAGGACCCTAC 720
              |||||||| |||| ||||||| ||| || |||||||| ||||| |||||| |
Db   661 AGGGCAGCTTGGGTGGGCTGGAGGAGCACTTCTCACACTCCTGCTCGCTCGGGCGGGCCCTAT 720

Qy   721 AACGCCTCCAACGTGGCCAGCTTCCTGTACCCCAATCTAGGAGGCTCCTGCGGAAGCTG 780
              |||| || | | ||||| |||| | ||||| |||| ||||| | || | || || |
Db   721 AATGCTTCCAATGTGGCTAGTTCATAAACCCGGACTTAGAAGGCTTCCTGACCGTTACTTGGGA 780

Qy   781 GGGCCATCACGCGGTGCTGGAGTGTGCTTAATCCGCTGGTGGACCGGTTACTTGGGA 840
              |||||||| | ||| ||||||| || ||| | ||| || || || || |||||||
Db   781 GGGCCATCACGAGGCCTGGAGTGTGGTGCTCAACCCACTGGTGACTGTCACTGCTGGA 840

Qy   841 AGGGGTCCTGCCTGAAGACAGTGTGCGGCAAGACGCAAGGGGCAAGTCCCAGAAG 900
              ||||||||||| ||| | || ||||||| || |||||| | ||||||||||||||
Db   841 ACAGGTCCTGGACAGGGACAATATGTGACAGGGACTCAAGAGGGACAATTCAGAAG 900

RESULT 13
AAV53631
ID   AAV53631 standard; cDNA; 1841 BP.
XX
AC   AAV53631;

Qy   78 CATCCGAGGCGGCGACGGCCCACGCCCGGCTCCGTCCGTTCACCCGTAGGGCTCTACGCCCT 137
         |||| | | | || | | |||| || ||||||||||||||| | ||||||||||
Db 394 CCTGGTTGGTCTTCGTGTGGGCAAGCTGCGAGCGCCGCCGTGGGCCGTGAAGGCGTGGAGGC 453

Qy 138 GAACCTGGGCTGCTCCGACCTGCTGACAGTCTCTGCCCCTGAAGGCGTGGAGGC 197
         |||| || | | || || | || || | | |||||||||||||||||||
Db 454 CAACCTGACGGCCTCGGACCTGCTCCTGCTGTTCCTGCCTTCCGCATGGTGGAGGC 513

Qy 198 GCTAGGCTCCGGGCGCTGGCCTCGCGCGTGCCCGTGCTTCGCGGCGGCCA 257
         || | |||||||| || || || | ||||||||||| ||||| || |
Db 514 AGCCAATGGCATGCACTGGCCCTGCCCTTCCATCCTCTGGATTCATCTT 573

Qy 258 CTTCTTCCCACTCTATGCCGGCGGGGGCTTCCTGGCCGCCCCTGAGTGCAGCGCCGCTACCT 317
         |||| || ||||| ||| | || || | ||| |||||
Db 574 CTTCACCACCATCTATCTCACCGCCCTTCCTGGCAGCTGTGAGCATTGAACGCTTCCT 633

Qy 318 GGGAGCAGCCTTCCCCTGGGCTACCAAGCCTTCCGAGGCCGTGCTATTCCTGGGGGGT 377
         |||||| ||| || ||||| ||| || | | || || | | ||||||| ||
Db 634 GAGTGTGGCCCACCCCCTGTGTACAGACCCGGCCGAGGCTGGGCTGGCAGGCAGGTCTGGT 693

Qy 378 GTGCGGCGCCATCTGGGCCCTCGGTCTGTGTCACCTGGGTCTCTTGGGTGTTGGGAGGC 437
         || ||| |||| ||| | | | | | || || ||| || ||||| ||||| | |
Db 694 GAGTGTGGCCCTGCTGTGTGGCCTCTGCTCACTGCAGCGTGGTCTCAGCGTCATAGAATT 753

Qy 438 TCCAGGAGGCTGGCTGGACCACAGCAACACCTCCCTGGGCATCAACACACCGGTCAACGG 497
         | |||||| | || | ||| ||| | ||| ||| | | | ||
Db 754 CTCAGGGG--------------ACATCTCCAGCAGCAGCAGCCAGGGACACCCAATGGGACCTG 795

Qy 498 CTCTCCGGTCTGCCTGGAGGCTGGACCCCGGCCGCGCCGCCGCGCCCGCTTCAGCCT 557
         ||||| | | | | | ||| | | | || | || | |||| | | | |
Db 796 CTACCTGGAGTTCCGGAAGGACCAG--CTAGCCATCCTCCTGCCGTGGCGGCTGGAG-AT 852

Qy 558 CTCTCCTCCTGCTCTTTTTCTGCCCTTGGCCATCAGCCCTTCTCACAGCCTTCTGCTACGTGGGCTGCCT 617
         |||||| ||||| | |||| ||| || || || || || |||||| | || || |||| |||||
Db 853 GGCTGTGGTCCTCTTCTTTGGTGTCCCGCTGATCATCACCAGCTACTGCTAC--AGCCCGCCT 909

Qy 618 CCGGGCACTGGCCGCCTCGGCGTCCCGGCTGACGCGACCAGCGGCAGAGTCGGCGGGCCGCCTGGGTGGC 677
         ||||||| | | | |||| | | | | ||| ||| | |||| | | |||| |||
Db 910 GGTGTGGATCCTCGGCAGAGGGGGCAGCACCCGGCCGCAGAGGAGGGTGGCGGGGGCTGTT 969

Qy 678 CGGGCGGGGCCCCTCCTCACGCTGCTGCTTCCTGGAGGACCCTACAAACGGCCTACAACGTGGC 737
         ||||| | || ||| || || || | | | ||| || || |||| || || ||||
Db 970 GGCGGCCACGCTGCTGCTCAACTTCCTTGTCTTCTGGGCCCTACAACGTGTCCCATGTCGT 1029

Qy 738 CAGCTTCCTGTAC 750
         ||||| ||||
Db 1030 GGGCTATATCTGC 1042

RESULT 14
AAA30762
ID   AAA30762   standard; cDNA; 1041 BP.
XX
AC   AAA30762;
XX
DT   21-AUG-2000 (first entry)
XX
DE   Human G protein-coupled receptor GPR41 cDNA.
XX
KW   G protein-coupled receptor; GPCR; constitutively active;
KW   intracellular loop 3; transmembrane domain 6; drug screening; agonist;
KW   antagonist; ss.
XX
OS   Homo sapiens.
XX
PN   WO20022129-A1.
XX
PD   20-APR-2000.
XX
PF   12-OCT-1999;   99WO-US023938.
XX
PR   13-OCT-1998;   98US-00170496.
XX

Qy   18 GCTCTCCTTCGGCCCTCTATGTGGGCCCCTTTGCCGCTGGGCTTCCCGCTGGGGCTTCCCGCTTCCGCTCCTGGC 77
         ||| ||||| ||| || | || | | ||| || || ||| |||| |||||||||| ||||||||||||
Db 334 GTTCGTCTTCTCGGTGTGTACCCTTCTCCACTTTCCTGGTGGCTCCCGGTGTGGGGGCTCCCCCCTCAACCTGCTGGCC 393

XX
DT   07-DEC-1998 (first entry)
XX
DE   Human 7-transmembrane receptor HNFDY20 cDNA.
XX
KW   HNFDY20; G-protein coupled receptor; human; infection; HIV; pain; cancer;
KW   anorexia; asthma; Parkinson's disease; acute heart failure; hypotension;
KW   hypertension; urinary retention; osteoporosis; angina pectoris;
KW   myocardial infarction; ulcer; allergy; benign prostatic hypertrophy;
KW   psychosis; anxiety; schizophrenia; manic depression; delirium; dementia;
KW   mental retardation; dyskinesia; Huntingdon's disease;
KW   Gilles de la Tourette's syndrome; therapy; diagnosis; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   125..1330
FT   /*tag= a
XX
PN   EP866126-A1.
XX
PD   23-SEP-1998.
XX
PF   16-FEB-1998;   98EP-00301122.
XX
PR   19-MAR-1997;   97US-00820521.
XX
PA   (SMIK)   SMITHKLINE BEECHAM CORP.
XX
PI   Bergsma DJ, Sathe GM, Fuetterer WS, Mao JY;
XX
DR   WPI; 1998-482962/42.
DR   P-PSDB; AAW59924.
XX
PT   New polynucleotides and polypeptides encoding a novel human 7-
PT   transmembrane receptor - useful for diagnosing and treating eg cancer,
PT   osteoporosis and Parkinson's disease and infections caused by HIV-1 or -
PT   2.
XX
PS   Claim 3; Page 2-4; 24pp; English.
XX
CC   This isolated polynucleotide codes for HNFDY20 (see AAW59924), a novel
CC   human 7-transmembrane G-protein coupled receptor. HNFDY20 polynucleotides
CC   may be obtained from cDNA libraries derived from mRNA in cells of human
CC   testis and human liver using expressed sequence tag analysis. The
CC   invention relates to HNFDY20 polypeptides and recombinant materials and
CC   methods for their production. It also provides methods for using such
CC   polypeptides and HNFDY20 polynucleotides for treatment of infections such
CC   as bacterial, fungal, protozoan and particularly HIV-1 or HIV-2
CC   infections, and conditions including pain, cancers, anorexia, asthma,
CC   Parkinson's disease, acute heart failure, hypotension, hypertension,
CC   urinary retention, osteoporosis, angina pectoris, myocardial infarction,
CC   ulcers, allergies, benign prostatic hypertrophy, and psychotic and
CC   neurological disorders, including anxiety, schizophrenia, manic
CC   depression, delirium, dementia, severe mental retardation and
CC   dyskinesias, such as Huntingdon's disease or Gilles de la Tourette's
CC   syndrome. Gene therapy using RNA encoding HNFDY20 can be used to treat
CC   conditions caused by under-expression of the protein. The invention also
CC   relates to methods of identifying agonists and antagonists and for using
CC   such compounds to treat conditions associated with HNFDY20 imbalance.
CC   Diagnostic assays for detecting diseases associated with inappropriate
CC   HNFDY20 activity or levels are also provided
XX
SQ   Sequence 1841 BP; 334 A; 538 C; 593 G; 376 T; 0 U; 0 Other;

Query Match   13.7%; Score 123.4; DB 2; Length 1841;
Best Local Similarity   52.3%; Pred. No. 4.7e-15;
Matches 383; Conservative   0; Mismatches 326; Indels   24; Gaps   4;

```
Qy  438 TCCAGGAGGCTGGCTGGACCACAGCAACACTCCCTGGCATCAACACACCGGTCAACGG 497
            ||||||  |||||||                ||||||||||||||||||||||||||
Db  465 CTCAGGGG------------ACATCTCCACAGCCAGCCAGGCACCAATGGGACCTG 506

Qy  498 CTCTCCGGTCTGCCTGGAGGCCTGGGACCCGGCCCTCTGCCGGCCCGGCTTCAGCCT 557
            |||||||||| |||||||  |||       ||||    ||||||||||||||||
Db  507 CTACCTGGAGTTCCGGAAGGACCAG--CTAGCACATCCTCCTGCCCGTGCGGCTGGAG-AT 563

Qy  558 CTCTCCTCGTCTTTTTCGCCCTTGGCCATCACAGCCTTCGCTGTACGTGGGCGTGCCT 617
            |||||||   |||  |  ||||||||||||||||||||   |||  |||||||||||
Db  564 GGCTGTGGTCCTCTTTGTGTGTCCGCTGATCATCACCAGCTACTGCTAC---AGCCGCCT 620

Qy  618 CCGGGCACTGGCCCGCCTCCGCCTGACGCGCACCAGCGGGAAGCTGCGGGCCCGGTGGC 677
            |||||     |||||||||||  | |||||||||||||||||||| |||  |||||
Db  621 GGTGTGGATCCTCCGCAGAGGGGCAGCAGCCGCCGCCGCAGGAGGTGGCGGGGCTGTT 680

Qy  678 CGGGGGGGCCCTCCTCACGCTGCTGCTGCTGGTAGGACCCTACCACCGCCTCCAACGTGGC 737
            |||||| ||||||||||| |||| |||   ||||| |||||||||||| |||||||||
Db  681 GGGCGGCCACGCTGCTCAACTTCCTGTCGTTGTCCTTTGGGCCCCTACAACGTGTCCATGTCGT 740

Qy  738 CAGCTTCCTGTAC 750
            ||| ||| |||
Db  741 GGGCTATATCTGC 753
```

RESULT 15
AAH26460

ID   AAH26460 standard; cDNA; 1041 BP.
XX
AC   AAH26460;
XX
DT   29-OCT-2001 (first entry)
XX
DE   Human G-protein coupled receptor GPR 42 cDNA.
XX
KW   GPR 42; G-protein coupled receptor 42; human; drug screening;
KW   dyslipidaemia; coronary heart disease; atherosclerosis; thrombosis;
KW   obesity; angina; kidney failure; peripheral vascular disease; stroke;
KW   diabetes; metabolic syndrome; syndrome X; antilipaemic; cardiant;
KW   antiatherosclerotic; thrombolytic; anorectic; antianginal; nephrotropic;
KW   vasotropic; cerebroprotective; antidiabetic; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO200161359-A2.
XX
PD   23-AUG-2001.
XX
PF   19-FEB-2001; 2001WO-GB000684.
XX
PR   18-FEB-2000; 2000GB-00003900.
PR   22-MAR-2000; 2000GB-00007015.
XX
PA   (GLAX) GLAXO GROUP LTD.
XX
PI   Wise A, Brown AJ;
XX
DR   WPI; 2001-536581/59.
DR   P-PSDB; AAB82759.
XX
PT   Identification of an agent for the modulation of G-protein coupled
PT   receptor useful for the treatment of disease i.e. dyslipidemia or stoke.
XX
PS   Claim 16; Page 46-47; 53pp; English.
XX
CC   The present sequence is that of cDNA encoding human G-protein coupled
CC   receptor 42 (GPR 42, see AAB82759). The invention is based on the finding
CC   that expression of GPR 42 and GPR 41 (see AAB82758) is restricted to
CC   adipose tissue. GPR 41 or GPR 42 may therefore be used as a screening
CC   target for the identification and development of novel pharmaceutical
CC   agents for use in inhibiting lipolysis. Methods are claimed for
CC   identifying agents that modulate GPR 41 or GPR 42 activity, involving:
CC   (1) contacting a test agent with GPR 41 or GPR 42 or a variant capable of

PA   (AREN-) ARENA PHARM INC.
XX
PI   Behan DP, Chalmers DT, Liaw CW;
XX
DR   WPI; 2000-329165/28.
DR   P-PSDB; AAY90680.
XX
PT   Non-endogenous constitutively activated human G protein-coupled
PT   receptors, useful for identifying agonists for use as pharmaceutical
PT   agents.
XX
PS   Example 1; Page 309-310; 341pp; English.
XX
CC   The invention relates to constitutively active, non-endogenous versions
CC   of endogenous human orphan G protein-coupled receptors (GPCRs, AAY90643-
CC   AAY90677 and AAY90683-Y90687), and to DNA encoding them (AAA30709-A30743
CC   and AAA30775-A30779). The mutant proteins of the invention contain a
CC   mutation in a portion of the protein comprising intracellular loop 3
CC   (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, X,
CC   is substituted for an endogenous residue in IC3 at a position 16 amino
CC   acids N-terminal of an endogenous proline in TM6 to form a sequence X-
CC   (AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg or
CC   Ala, and is preferably Lys. When the endogenous residue at this position
CC   is Lys, this residue is replaced by His, Arg or preferably Ala. The 15
CC   amino acid stretch between the substituted amino acid and the Pro may be
CC   endogenous, non-endogenous, or a mixture of endogenous and non-endogenous
CC   residues. The constitutively active GPCRs are useful for identifying
CC   antagonists, agonists and partial agonists for use as pharmaceutical
CC   agents. The mutant proteins are also useful in research settings for
CC   elucidating the roles of the receptors in normal and diseased conditions.
CC   Antagonists for a particular GPCR are useful for treating diseases and
CC   disorders associated with that receptor. Because the novel mutant GPCRs
CC   are constitutively active, they can be used directly for screening of
CC   compounds without the need for endogenous ligands. The present sequence
CC   represents cDNA encoding a human wild-type GPCR used in an
CC   exemplification of the invention. This was cloned and subjected to site-
CC   directed mutagenesis (SDM) to generate DNA encoding the corresponding
CC   mutant of the invention
XX
SQ   Sequence 1041 BP; 170 A; 324 C; 315 G; 232 T; 0 U; 0 Other;

Query Match   13.5%; Score 121.8; DB 3; Length 1041;
Best Local Similarity   52.1%; Pred. No. 9.4e-15;
Matches   382; Conservative   0; Mismatches 327; Indels   24; Gaps   4;

```
Qy   18 GCTCTCCTTCGGCCCTCTATGTGGCGCGCCTTTGCCGTCGGGCTTCCCGCTCAACGTCCTGGC 77
            ||||   |  ||||| |||  |||| ||| || ||||||| ||||| |||||||||||||||
Db   45 GTTCGTCTTCTCGGTGTACCTTCTCACTTTCCTGGTGGGCTCCCCCTCAACCTGCTCGC 104

Qy   78 CATCCGAGGCGACAGGCGCCACGCCCGCCGCCGGCTCCGTCTCACCCCTAGCGTCCCCT 137
            | |||  ||  | |  |||  || ||   ||||||| |||||  ||||||||||||||
Db  105 CCTGGTGGTCTTCGTGGGCAAGTCGAGCGTCCAGCGCCGCCGGTGGCCGTGGGCCGTCCTGCT 164

Qy  138 GAACCTGGGCGTGCTCCGACCTGCTGCTCGACAGTCTCTCGCCCCTGAAGGCGGTGGAGGC 197
            ||||||| || ||||| |||||||| |||          |||||||||||||||||||
Db  165 CAACCTGACCGGCTCGGACCTGCTCCTGCTGTTCCTGCCCTTTCCGCATGGTGGAGGC 224

Qy  198 GCTAGCCTCCGGGGGCCTCCGGGGCCTCGCGGCCTCGCGTGTCCCCCTCTTCTGCGTGGCCCA 257
            |||||||   || || ||||| || |||  ||| |||| |||   ||||||||  ||||
Db  225 AGCCAATGGCATGCACTGCGCCCTGCCCTTCATCCTCTGCCCAACTTCTCTGGATTCATCTT 284

Qy  258 CTTCTTCCCACTCTATGCCGGCGGGGGGCTTCCTGGCCGCCGCCCTGAGTGCAGGCGCTACCT 317
            |||| || ||| | || | ||| || ||  || ||||||||||| |||||| ||||| |||
Db  285 CTTCACCACCATCATCTCACCGCCCTCTTCCTGGCAGTCTGTGAGCATTGAAGCCTTCCT 344

Qy  318 GGGAGCAGCAGCCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCGTGCTATTCCTGGGGGGT 377
            ||| || ||||| ||| |  ||| |||  | |  |||||||||   || |||  |||||||
Db  345 GAGTGTGGCCCACCCACTGTGGTACAAGAACCCGGCCGAGGCTGGGGCAGGCAGGTCTGT 404

Qy  378 GTGCGCGGCCATCGGGCCCCTGCTCCGTGTCACCTGGGGTCGTGCTTTTGGGTTGGAGGC 437
            |||| || |  ||||| |  ||| |   ||||| ||| || ||||| ||||| |||||||
Db  405 GAGTGTGGCCTGCTGCTGTTGGCCTGTGCTGCTCGTCACGGCTGTTCTACGTGTCATAGAATT 464
```

```
CC   coupling to a G-protein; and (2) monitoring for GPR 41 or GPR 42 activity
CC   in the presence of a G-protein, thereby determining whether the test
CC   agent modulates activity. Such agents, which may be an activator of GPR
CC   41 or GPR 42, an inhibitor of lipolysis, or a polynucleotide encoding GPR
CC   41, GPR 42 or variant polypeptide, are useful for the treatment of
CC   dyslipidaemia, coronary heart disease, atherosclerosis, thrombosis or
CC   obesity, angina, chronic renal failure, peripheral vascular disease,
CC   stroke, type II diabetes or metabolic syndrome (syndrome X) (all claimed)
XX
SQ   Sequence 1041 BP; 169 A; 321 C; 317 G; 234 T; 0 U; 0 Other;

Query Match         13.5%;  Score 121.8;  DB 4;  Length 1041;
Best Local Similarity   52.1%;  Pred. No. 9.4e-15;
Matches  382;    Conservative   0;  Mismatches 327;  Indels  24;  Gaps    4;

Qy         18 GCTCTCCTTCGGCCTCTATGTGGCCGCCTTTGCCGCTGGGCTTCCCGCTGGGCTTCAACGTCCTGGC 77
              |||| ||| ||||    ||||| |||||||||||||||||  ||||||| |||||||
Db         45 GTTCGTCTTCGGTGTACTTTCTCACTTTCCTGTGGGGCTCCCCTCAACCTTGCTGGC 104

Qy         78 CATCCGAGGCGACAGGCCCACGCCCGGCTCCGTCTCACCCCTAGCCTGGTCTACGCCCT 137
              ||||| ||||   |||| ||||| |||| ||||||  ||| |||||||||
Db        105 CCTGGTGGTTCGTGTGGGCAAGCTGCGGTGCCGCCCGTGGCCGTGGACGTGCTCCTGCT 164

Qy        138 GAACCTGGGCTGCTCCGACCTGCTGCTGACAGTCTCTCTGCCCCTGAAGGCGGTGGAGGC 197
              |||||| |||||      |||| |||| |||||       ||||||| ||||||||
Db        165 CAACCTGACCGCCTCGGACCTGCTCCTGGTCGTCCTGCCTTTCCGCATGGTGGAGGC 224

Qy        198 GCTAGCCTCCGGGGCCTCTGCCTCCGGCCTCGGCTCCTCGGGCTCGTCTTCGCGGTGGCCCA 257
              ||||| || ||||| ||  ||||||| |||| |||||| |||||
Db        225 AGCCAATGGCGATGCACTGGCCCCCTGCCCCTTCATCCTCTGCCCACTCTCTGGATTCATCTT 284

Qy        258 CTTCTTCCCACTCTATGCCGGCGGGGGCTTCCTGGCCGCCCTGAGTGCAGGCCGCTACCT 317
              |||||     ||||  || |||||| |||||||| |||||||
Db        285 CTTCACCACCATCTATCTCACCGCCCGTTCTTCCTGCAGCTGTGAGCATTGAACGCTTCCT 344

Qy        318 GGGAGCAGGCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCGTGCTATTCCTGGGGGGT 377
              |||||||  |||||||||||||     |||||||| |  ||||| |||||||||
Db        345 GAGTGTGGCCCACCCACTGTGGTACAAGACACCGGCCGAGGCTGGGGCAGGTCTGGT 404

Qy        378 GTGCGGCGGCCATCTGGGCCCCTGCCTCGTGTGTCACCTGGGTCGTCTTTGGGTTGGAGGC 437
              ||| ||||  ||||   |  |||  |||||||  |||||||| ||||||
Db        405 GAGTGTGGCCTGCTGGCCTCTGCTCACTGCAGCGGTGGCTCACGTGTCATAGAATT 464

Qy        438 TCCAGGAGGCTGGCTGGACCACAGCAACACCTCCCTGGGCATCAACACACACCGGTCAACGG 497
              |||||||
Db        465 CTCAGGGG------------------ACATCTCCCACAGCCAGGGCACCCAATGGGACCTG 506

Qy        498 CTCTCCGGTCGTCGCCTGGAGGCCTGGGACCCGGGCCCGGCCCGGCCGCCCTTCAGCCT 557
              ||||||  ||||||||| |||||       |||| ||||||||||||
Db        507 CTACCTGGGAGTTCTGGAAGGACCAG--CTAGCCATCCTCGTGCCCGGCTGGGAGA-T 563

Qy        558 CTCTCCTGCTCTTTTTCTCCGCCCTGGCCATCACAGCCTTCTGCTACGTGGGCTGCCT 617
              |||||||||||||  ||||||||| |||||||| ||||||
Db        564 GGCGTGGTGTCCTCTTCTGTGTCCCGCTGATCATCACCAGCTACTGGTCACAGCCGCCTGGT 623

Qy        618 CCGGGCACTGGCCCCGTCCGGCTCCGGCCCTGACGCACAGAGCGGAGCTGCCGGCCCTGGGTGGC 677
              |||||||||||||||     |||||||||| ||||||   ||||| ||| |||||||||
Db        624 GTGGATCCTCCGGCAGGGGGGC---AGCCACCGCCCGGCAGGAGGAGGGGTGCGGGGGCTGGT 680

Qy        678 CGGCGGGCCCTCCTCACGCTGCTGCTCTTGCGTAGGACCCTACAAACGCTGGC 737
              |||||      ||||  |||||  ||||||||
Db        681 GGCGGCCACGCTGCTCAACTTCCTTGTCTGCTTTGGGCCCTACAACGTGTCCCATGTCGT 740

Qy        738 CAGCTTCCTGTAC 750
              ||| ||||||
Db        741 GGGCTATATCTGC 753
```

Search completed: April 30, 2004, 15:41:49
Job time : 456 secs

THIS PAGE BLANK (USPTO)

```
                    Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:        April 30, 2004, 14:33:11 ; Search time 2948 Seconds
                                           (without alignments)
                                         9147.065 Million cell updates/sec

Title:          US-10-202-687-1
Perfect score:  903
Sequence:       1 atggacctgccccccagct.........gggcaagtcccagaagtaa 903

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       27513289 seqs, 1493109276 residues

Total number of hits satisfying chosen parameters:   55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  EST:*
        1:   em_estba:*
        2:   em_esthum:*
        3:   em_estin:*
        4:   em_estmu:*
        5:   em_estov:*
        6:   em_estpl:*
        7:   em_estro:*
        8:   em_htc:*
        9:   gb_est1:*
       10:   gb_est2:*
       11:   gb_htc:*
       12:   gb_est3:*
       13:   gb_est4:*
       14:   gb_est5:*
       15:   em_estfun:*
       16:   em_estom:*
       17:   em_gss_hum:*
       18:   em_gss_inv:*
       19:   em_gss_pln:*
       20:   em_gss_vrt:*
       21:   em_gss_fun:*
       22:   em_gss_mam:*
       23:   em_gss_mus:*
       24:   em_gss_pro:*
       25:   em_gss_rod:*
       26:   em_gss_phg:*
       27:   em_gss_vrl:*
       28:   gb_gssl:*
       29:   gb_gss2:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	Match	ID	Description
1	901.4	99.8	903	29	AY411510	AY411510	Homo sapi
2	823.6	91.2	834	29	AY411511	AY411511	Pan trogl
3	602.8	66.8	903	29	AY411512	AY411512	Mus muscu
4	376.8	41.7	443	10	AW583167	AW583167	iaiifil.y

```
 5    306   33.9   522   29   CE730455   CE730455 tigr-gss-
 6  295.2   32.7   609   14   CA776599   CA776599 ip08c01.y
 7  189.8   21.0   234   29   CE038975   CE038975 tigr-gss-
 8  132.4   14.7   961   29   CNS041EX   AL270114 Tetraodon
 9    130   14.4   774   28   BZ160914   BZ160914 CH230-339
10    128   14.2   543   28   AQ119988   AQ119988 HS 3024 A
11  117.4   13.0  1117   12   BM544968   BM544968 AGENCOURT
12  106.6   11.8   779   14   CD559956   CD559956 AGENCOURT
13  103.6   11.5   696   14   CF147780   CF147780 AGENCOURT
14  102.8   11.4   472   12   BI541693   BI541693 455527 MA
15   95.8   10.6   489   14   CA777066   CA777066 ip02f08.y
16   83.8    9.3   619   28   AZ082043   AZ082043 RPCI-23-2
17   83.8    9.3   672   13   BY741744   BY741744 BY741744
18   81.2    9.0   757   12   BG190636   BG190636 Mus muscu
19   80.2    8.9  1077   29   AY403859   AY403859 RST9710 A
20     80    8.9   738   14   CD559659   CD559659 AGENCOURT
21     80    8.9   745   14   CD559660   CD559660 AGENCOURT
22     80    8.9   766   14   CD559663   CD559663 AGENCOURT
23     80    8.9   772   14   CD559662   CD559662 AGENCOURT
24     80    8.5  1098   29   AY408183   AY408183 Homo sapi
25   76.8    8.5   786   29   AY408184   AY408184 Pan trogl
26   73.6    8.2   681   14   CF365470   CF365470 836101 MA
27   73.2    8.1  2297   11   AK078861   AK078861 Mus muscu
28   71.6    7.9  1257   11   AK050524   AK050524 Mus muscu
29   71.2    7.9   504   10   AW632495   AW632495 92464 MAR
30   70.2    7.8   500   10   BF603623   BF603623 269181 MA
31     70    7.8   557   13   BQ807875   BQ807875 NISC_kk11
32     70    7.8  1046   29   AY407666   AY407666 Mus muscu
33     70    7.8  2323   11   BC009877   BC009877 Homo sapi
34   69.6    7.7   631   11   CB499474   CB499474 ssalkhc01
35   68.8    7.6  2912   11   AK087410   AK087410 Mus muscu
36   68.6    7.6   460   10   AW206493   AW206493 UI-H-BI1-
37   68.6    7.6   462   10   AW138148   AW138148 UI-H-BI1-
38   68.6    7.6   544    9   AI344017   AI344017 tc01e02.x
39   68.4    7.6   699   10   BE264355   BE264355 60119141
40   68.4    7.6   843   29   CNS00CS1   AL059666 Drosophil
41   68.4    7.6   925   29   CNS0091P   AL053013 Drosophil
42   68.4    7.6  4435   11   AK049671   AK049671 Mus muscu
43   68.2    7.6   916   29   CNS02BZX   AL194406 Tetraodon
44   67.2    7.4   553   12   BM141985   BM141985 if25b05.y
45   67.2    7.4   611   28   AZ409268   AZ409268 1M0180F17
```

                                                    ALIGNMENTS

```
RESULT 1
AY411510
LOCUS           AY411510               903 bp    DNA      linear    GSS 16-DEC-2003
DEFINITION      Homo sapiens GPR40 gene, VIRTUAL TRANSCRIPT, partial sequence,
                genomic survey sequence.
ACCESSION       AY411510
VERSION         AY411510.1  GI:39767478
KEYWORDS        GSS.
SOURCE          Homo sapiens (human)
  ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1  (bases 1 to 903)
  AUTHORS       Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
                Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
                Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
                Adams,M.D. and Cargill,M.
  TITLE         Inferring nonneutral evolution from human-chimp-mouse orthologous
                gene trios
  JOURNAL       Science 302 (5652), 1960-1963 (2003)
   PUBMED       14671302
REFERENCE       2  (bases 1 to 903)
  AUTHORS       Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
                Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
                Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
                Adams,M.D. and Cargill,M.
  TITLE         Direct Submission
```

JOURNAL   Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
          Rockville, MD 20850, USA
COMMENT   This sequence was made by sequencing genomic exons and ordering
          them based on alignment.
FEATURES   Location/Qualifiers
     source   1..903
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     gene   <1..>903
                     /gene="GPR40"
                     /locus_tag="HCM4248"
ORIGIN

Query Match   99.8%; Score 901.4; DB 29; Length 903;
Best Local Similarity   99.9%; Pred. No. 2.2e-148;
Matches 902; Conservative   0; Mismatches   0; Indels   1; Gaps   0;

```
Qy    1 ATGGACCTGCCCCCCAGCTCTCCTTCGGCCTCTATGTGCCGCCTTTGCGCTGGCTTC 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGGACCTGCCCCCCCAGCTCTCCTTCGGCCTCTATGTGCCGCCTTTGCGCTGGCTTC 60

Qy   61 CCGCTCAACGTCCTGGCCATCCGAGGGCGCGACCGGCCCAGCCCGGCTCCGTCTCCACCCCT 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CCGCTCAACGTCCTGGCCATCCGAGGCGCGACGGCCGCCCAGCCCGGCTCCGTCTCCACCCCT 120

Qy  121 AGCCTGGTCTACGGCCCTGAACCTGGGCTGCTCCAGCAGTCCTCGGGGGCCCTGCCC 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 AGCCTGGTCTACGGCCCTGAACCTGGGCTGCTCCGACCTGCTGCTGACAGTCTCTCTGCCC 180

Qy  181 CTGAAGGGCGGTGGAGGCGCGCTAGCCTCCGGGGGCTTCCTGGCCCTCGTGTGCCCC 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 CTGAAGGGCGGTGGAGGCGCGCTAGCCTCCGGGGGCTTCCTGGCCCTCGTGTGCCCC 240

Qy  241 GTCTTCGCGGTGGCCCCACTTCTTCCCACTCTATGCCGGGCGGGGCTTCCTGGCCCCCCG 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 GTCTTCGCGGTGGCCCCACTTCTTCCCACTCTATGCCGGGCGGGGCTTCCTGGCCCCCCG 300

Qy  301 AGTGCAGGCCGCTACCTGGGAGGCAGCAGCCTTCCCCCTTGGGCTACCAAGCCTTCCGGAGGCCG 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 AGTGCAGGCCGCTACCTGGGAGGCAGCAGCCTTCCCCCTTGGGCTACCAAGCCTTCCGGAGGCCG 360

Qy  361 TGCTATTCCTGGGGGTGTGCGCGCCATCGGCCTCCGTCTGTCACCTGGTCTG 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 TGCTATTCCTGGGGGTGTGCGCGCCATCGGCCTCCGTCTGTCACCTGGTCTG 420

Qy  421 GTCTTTGGGTTGGAGGCTCCAGGAGGGCTGCTGACCACAGCAACACCTCCCTGGCATC 480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 GTCTTTGGGTTGGAGGCTCCAGGAGGGCTGCTGACCACAGCAACACCTCCCTGGCATC 480

Qy  481 AACACACCGGTCAACGGCTCTCCGGTCTCGTCGCTGCCTCGGAGGCCTCGGCCGC 540
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 AACACACCGGTCAACGGCTCTCCGGTCTCGTCGCTGCCTCGGAGGCCTCGGCCGGC 540

Qy  541 CCGGCCGCTTCAGCCTTTCTGCCCTCTTTTTCTGCGCTCCGGCCCATCACAGCCTTC 600
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 CCGGCCGCTTCAGCCTTTCTGCCCTCTTTTTCTGCGCTCCGGCCCATCACAGCCTTC 600

Qy  601 TGCTACGTGGGGCTGCCTCCGGGCACTCGGCCCGGCCTGACGGCACAGGCGGAAGCTG 660
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 TGCTACGTGGGGCTGCCTCCGGGCACTGGGGCCCACTCGGCCCGGCCTGACGGCACAGGCGGAAGCTG 660

Qy  661 CGGGGCCGGCCTGGGTGGCCGGGGCGGGGCCTCCACGCTGCTCGTCTCGGGACCCTAC 720
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 CGGGGCCGGCCTGGGTGGCCGGGGCGGGGCCTCCACGCTGCTCGTCTCGGGACCCTAC 720

Qy  721 AACGGCTCCAACGTGGCCAGCTTCCTGTACCCCAATCTGGCGGAGGCTCCTGGCCGGAAGCTG 780
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 AACGGCTCCAACGTGGCCAGCTTCCTGTACCCCAATCTGGCGGAGGCTCCTGGCCGGAAGCTG 780

Qy  781 GGGGCTCATCACGGGTGCCTGGAGTGTGGTGCTTAATCCGCTGGTTACTTGGGA 840
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 GGGGCTCATCACGGGTGCCTGGAGTGTGGTGCTTAATCCGCTGGTTACTTGGGA 840

Qy  841 AGGGGTCCTGGCTGGAAGACAGTGTGTGGGCAAGAACGCAAGGGGGCAAGTCCCAGAAG 900
         ||||||||||||| ||||||||| ||||||| ||||||||||||||||||||||||||||
Db  841 AGGGGTCCTGGCCTGAAGACACAGTGTGTGCGGCAAGAACGCAAGGGGCAAGTCCCAGAAG 900

Qy  901 TAA 903
         |||
Db  901 TAA 903
```

RESULT 2
AY411511
LOCUS   AY411511   834 bp   DNA   linear   GSS 16-DEC-2003
DEFINITION Pan troglodytes GPR40 gene, VIRTUAL TRANSCRIPT, partial sequence,
            genomic survey sequence.
ACCESSION   AY411511
VERSION   AY411511.1 GI:39767479
KEYWORDS   GSS.
SOURCE   Pan troglodytes (chimpanzee)
  ORGANISM Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1 (bases 1 to 834)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE   Inferring nonneutral evolution from human-chimp-mouse orthologous
            gene trios
  JOURNAL   Science 302 (5652), 1960-1963 (2003)
   PUBMED   14471302
REFERENCE   2 (bases 1 to 834)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT   This sequence was made by sequencing genomic exons and ordering
            them based on alignment.
FEATURES   Location/Qualifiers
     source   1..834
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
     gene   <1..>834
                     /gene="GPR40"
                     /locus_tag="HCM4248"
ORIGIN

Query Match   91.2%; Score 823.6; DB 29; Length 834;
Best Local Similarity   99.0%; Pred. No. 9.9e-135;
Matches 826; Conservative   0; Mismatches   8; Indels   0; Gaps   0;

```
Qy   70 GTCCTGGCCATCCGAGGGCGCGACCGGCCCAGCCCGGCTCCGTCTCCACCCCTAGCCTGGTC 129
         ||||| ||||||||||| ||| ||| ||||| |||||||||||||||||||||||||||||
Db    1 GTCCTNNCCATCCGAGGCGCNNGACGGCCANGCCCGGCTCCGTCTCCACCCCTAGCCTGGTC 60

Qy  130 TACGGCCCTGAACCTGGGCTGCTCCAGCAGTCCTCGGGGGCCCTGAAGGCG 189
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 TACGGCCCTGAACCTGGGCTGCTCCGACCTGCTGCTGACAGTCTCTCTGCCCCTGAAGGCG 120

Qy  190 GTGGAGGCGCGCTAGCCTCCGGGGGCTTCCTGGCCCTCGTGTGCCGTCTTCGCG 249
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 GTGGAGGCGCGCTAGCCTCCGGGGGCTTCCTGGCCCTCGTGTGCCGTCTTCGCG 180

Qy  250 GTGGCCCCACTTCTTCCCACTCTATGCCGGGCGGGGCTTCCTGGCCCTCGTGAGTCGCAGGC 309
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 GTGGCCCCACTTCTTCCCACTCTATGCCGGGCGGGGCTTCCTGGCCCTCGTGAGTCGCAGGC 240

Qy  310 CGCTACCTGGGACGAGCAGCCTTCCCCCTTGGGCTACCAAGCCTTCCGGAGGCCGTGCTATTCC 369
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  310
```

```
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10090"
     gene           <1..>903
                    /gene="GPR40"
                    /locus_tag="HCM4248"

ORIGIN

Query Match     66.8%;  Score 602.8; DB 29;  Length 903;
Best Local Similarity 79.3%;  Pred. No. 6.3e-96;
Matches 715;  Mismatches 187;  Indels 0;  Gaps 0;
             Conservative 0;

Qy    1  ATGGACCTGCCCCCGCAGCTCTCCTTGGCGCCTTTGCGCTGGGGCTTC   60
         |||||||||||||  ||||||||||||  ||  ||||||||||||||||
Db    1  ATGGACCTGCCCCCACAGCTCTCCTTCCTCTATGTATCTGCTTGGCGCTGGGGCTTT   60

Qy   61  CCGCTCAACGTCCTGGCCATCAACAGCCTCCGAGGCGGACGGCCACGCGCCTCGTCTCACCCCT  120
         ||  |  ||||  || | ||  ||  | |||||||||||  |||  ||||||  |   |||||||
Db   61  CCATTGAACTTGTTAGCGCTCACCAGCAGTGTCCCAGCTAAACTGGACTCACTCCC  120

Qy  121  AGCCTGGTCTACGCCCTGACCTGGGCTGCTCCGACCTGCTGTGTCCGACAGTCTTCTGCCC  180
         |||  ||||||| ||  | | | ||| |||||| ||  ||  | ||  | | || |||||||
Db  121  AGCTTGGTCTACACTCTCCATCTGGGCTGCTCGATCTCGATCTTCCTACTGGCCATCACTCTGCCC  180

Qy  181  CTGAAGGCGGTGGAGGGCTAGCCTCCGGGGGCTAGGACCCTACAAGCGCTCC  240
         ||||||||||||  |||  ||  |||| |||||  ||| |   ||| ||||||
Db  181  CTGAAGGCTGTGTGGAGCCCTGGCTTCTGGAGCTGTACGAGGGCGTGTGCCCTG  240

Qy  241  GTCTTCGCGGTGGCCCACTTCTTCCCACTCTCCTCTCCCCTGGGGGCCGGGGCGGCCG  300
         ||||||||||||||| || ||||| |||  ||||| ||||| |||||  |||||||||
Db  241  GTCTTGCCTTGGCCCCACTTTGCTCCACGGGGAGCGGAGGCTCCTAGCCTGGTCTC  300

Qy  301  AGTGCAGGCCGTACCTGGGAGCAGCCTTCCCCTTGGGCTACCAAGGCTTCCGGAGGCCG  360
         |||  ||| ||  || |||||||||| ||||| |||||||| |||||  ||||||  ||
Db  301  AGCGCTGGCCGTACCTGGGAGGCCGGAGGCCGCATCCCGGGTACCATCCGGGGGCCC  360

Qy  361  TGCTATTCCTGGGGGTGTGCGGGCCCTGTCTTCGTCACCTGGGTCTG  420
         ||  ||||| ||||||||||||||| | || | |||| || |||  |
Db  361  CGCTATTCCTGGGGGTGTGCGGGCCCTGTATATGGGCCTCTGCTGCCACCTGGGGCATC  420

Qy  421  GTCTTTGGGTTGGAGGCTCCAGGAGGCTGCTGAACAGCAGCCTGGGTGCCGGCATC  480
         || ||||||  || | |   || |||||  |||||||| | |  |   ||||||||
Db  421  GCCCTTGGCTTGGAGACTTCCGGAAGCTGGTGGAACAGCTACCAGTTCCCTGCGGCATC  480

Qy  481  AACACACCGGTCAACGGCTTCGTCCGGTCGCTGCTCCGAGGCCTGGGACCCCTGCCGGC  540
         ||  ||  | |||  ||||  ||  ||  |||  || ||| || ||| |   || |||||
Db  481  AACATACCGTGAATGGCTCGTCCCGGTGCTGCGCTGGTGGATCCCGGCGTCCCGACTCTCGCGCCGC  540

Qy  541  CCGGCCGCTTGGGTGTCAGCCTCCAGCCGCCATCACAGCCTTC  600
         || ||| | || |||  | || |||   ||| |||| |||  |
Db  541  CCTGCCCGTCTCAGTTTCTCCATTCTGTCTCTTCTTGGCTCATCACTGCCCTTC  600

Qy  601  TGCTACGTGGGCGCTGCTCGGGGCCGGGGCCACTGGTGTCGTCGTAGGACCCTAC  660
         || ||| || |||  || |||||| ||| |||  |||||  ||| ||||| |||
Db  601  TGCTATGTGGGCTGCTCGGGGCCGGGGGGCCGGAGCCGAGGCCACAAACGGAAGCTC  660

Qy  661  CCGGGCCGTGGGTGGCCGGCGGGGCCCTCTCCTCACGGTGCTCTGCGTAGGACCCTAT  720
         || |||||  || || |||  ||||||| ||||| ||  || |||| | ||||| |||
Db  661  AGGGCAGCTTGGGTGGCCGGAGGGGCCGCTCTCTTCTTCGCTCTGCTTGGTCATCACTGCCCTTC  720

Qy  721  AACGGCTCCAACGTGCCAGCTGCCGGCGGAAGCTG  780
         ||||| || |||  |  | |   ||||||| ||||
Db  721  AATGCCTCCAATGTGGCTAGTTTCATAAACCCGGAAGCTCTAGGAGGCTCCTGGGAAGAGTTA  780

Qy  781  GGGGTCATCACGGGTGCCTGGAGTGGGTGCTTAATCCGGTTACTTGGGA  840
         ||| ||||||| || |  ||| || | || || ||||| ||||||||||
Db  781  GGACTCATCATCACAGCGGTAGTGGTGGTACTCAACCCCACTACTTGGGA  840

Qy  841  AGGGGTCCTGCCTGAAGACAGTGTCGCGGCAAGAACGCAAGGGGCAAGTCCCAGAAG  900
         |||||| |  | | |||||||||| |  || ||| | |  || |||||| ||||||||
Db  841  ACAGGTCCTGACGCGGGAACAATATGTGTGACGAGGACTCAAAGAGAGAACAATTCAGAAG  900

Qy  901  TA  902
         ||
Db  901  ...
```

RESULT 3

AY411512

```
LOCUS       AY411512          903 bp    DNA     linear   GSS 16-DEC-2003
DEFINITION  Mus musculus GPR40 gene, VIRTUAL TRANSCRIPT, partial sequence,
            genomic survey sequence.
ACCESSION   AY411512
VERSION     AY411512.1  GI:39767480
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 903)
  AUTHORS   Clark,A.G., Glanowski,S., Nielsen,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Inferring nonneutral evolution from human-chimp-mouse orthologous
            gene trios
  JOURNAL   Science 302 (5652), 1960-1963 (2003)
   PUBMED   14671302
REFERENCE   2  (bases 1 to 903)
  AUTHORS   Clark,A.G., Glanowski,S., Nielsen,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT     This sequence was made by sequencing genomic exons and ordering
            them based on alignment.
FEATURES             Location/Qualifiers
     source          1..903
                     /organism="Mus musculus"
```

```
Db     901 TA 902

RESULT 4
AW583167
LOCUS       AW583167                  443 bp    mRNA        linear   EST 13-MAR-2002
DEFINITION  ia1lfi1.y1 Human Pancreatic Islets Homo sapiens cDNA clone
            IMAGE:563029 5' similar to SW:GP40 HUMAN O14842 PUTATIVE G
            PROTEIN-COUPLED RECEPTOR GPR40. [1] ;, mRNA sequence.
ACCESSION   AW583167
VERSION     AW583167.1  GI:7260111
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 443)
  AUTHORS   Melton,D., Meadows,A., Clifton,S., Hillier,L., Marra,M., Pape,D.,
            Wylie,T., Martin,J., Bliskain,A., Schmitt,A., Theising,B.,
            Ritter,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M.,
            McCann,R., Cole,R., Tsagareishvili,R., Williams,T., Jackson,Y. and
            Bowers,Y.
  TITLE     WashU-Harvard Pancreas EST Project
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Libraries were constructed by Dr. Douglas Melton
            DNA sequencing by: Washington University Genome Sequencing Center
            For information on obtaining a clone please contact: Juliana Brown
            (brown@fas.harvard.edu)
            This sequence now available from the IMAGE consortium, for clone
            orders contact: info@image.llnl.gov
            Seq primer: -40RP from Gibco
            High quality sequence stop: 421.
FEATURES             Location/Qualifiers
     source          1..443
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5638029"
                     /tissue_type="Islets of Langerhans"
                     /dev_stage="Adult"
                     /lab_host="DH10B"
                     /clone_lib="Human Pancreatic Islets"
                     /note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
                     Site_2: Sal 1; Library constructed using SuperScript
                     Plasmid Library kit (Life Technologies). cDNA made by
                     oligo-dT priming. Size-selected by column fractionation;
                     average insert size 1.08 kb. Primary library,
                     unamplified."

ORIGIN

Query Match        41.7%;   Score 376.8; DB 10; Length 443;
Best Local Similarity   99.5%;  Pred. No. 2.6e-56;
Matches  378; Conservative    0; Mismatches    2; Indels    0; Gaps     0;

Qy   524 ACCCGGCCTCTGCGGCCCGGCCCGGCCTTCAGCCTCTCTCTCCGCTCTTTTTTCTGCCCT 583
         |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db     1 ACCCGGCCTCTGCGGCCCGGCCCGGCCTTCAGCCTCTCTCTCCGCTCTTTTTTCTGCCCT 60

Qy   584 TGGCCATCACAGCCTTCTGCTACGTGGGCTGCCTCCGGGCACTGGCCCCGCCTCCGGCTGA 643
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 TGGCCATCACAGCCTTCTGCTACGTGGGCTGCCTCCGGGCACTGGCCCCACTCCGGCCTGA 120

Qy   644 CGCACAGGCGGAAGCTGCGGGGGCCGCCTGGGTGGCCGGCGGCGGGGCCCTTCCTCCACCGCTGTGC 703
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CGCACAGGCGGAAGCTGCGGGGGCCGCCTGGGTGGCCGGCGGCGGGGCCCTTCCTCCACCGCTGTGC 180
```

```
Db     901 TA 902

Qy   704 TCTGCGTAGGACCCTACACACGCCTTCCTGTACCCAATCTAGGAG 763
         |||||||||||||||||||||||||||||||||||||||||||||
Db   181 TCTGCGTAGGACCCTACACACGCCTTCCTGTACCCAATCTAAGAG 240

Qy   764 GCTCCTGGCGGAAGCTGGGGCTCATCACGGGTGGTGTGGCTTAATCCGCTGG 823
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GCTCCTGGCGGAAGCTGGGGCTCATCACGGGTGGTGTGGCTTAATCCGCTGG 300

Qy   824 TGACCGGTTACTTGGGAAGGGGTCCTGGCCTGAAGACAGTGTGCGGCAAGAACGCAAG 883
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 TGACCGGTTACTTGGGAAGGGGTCCTGGCCTGAAGACAGTGTGCGGCAAGAACGCAAG 360

Qy   884 GGGGCAAGTCCCAGAAGTAA 903
         ||||||||||||||||||||
Db   361 GGGGCAAGTCCCAGAAGTAA 380

RESULT 5
CE730455
LOCUS       CE730455                  522 bp    DNA        linear   GSS 30-SEP-2003
DEFINITION  tigr-gss-dog-17000315416692 Dog Library Canis familiaris genomic,
            genomic survey sequence.
ACCESSION   CE730455
VERSION     CE730455.1  GI:37070575
KEYWORDS    GSS.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1  (bases 1 to 522)
  AUTHORS   Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
            Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
            Venter,J.C.
  TITLE     The dog genome: survey sequencing and comparative analysis
  JOURNAL   Science 301 (5641), 1898-1903 (2003)
   MEDLINE  22875432
   PUBMED   14512627
COMMENT     Contact: Kirkness EF
            The Institute for Genomic Research
            Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
            Rockville, MD 20850, USA
            Tel: 301-838-0200
            Fax: 301-838-0208
            Email: ekirknes@tigr.org
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1..522
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /strain="Standard Poodle"
                     /db_xref="taxon:9615"
                     /clone_lib="Dog Library"
                     /note="Site_1: BstXI; Libraries were prepared from
                     peripheral blood"

ORIGIN

Query Match        33.9%;   Score 306; DB 29; Length 522;
Best Local Similarity   86.0%;  Pred. No. 7.6e-44;
Matches  339; Conservative    0; Mismatches   55; Indels    0; Gaps     0;

Qy     1 ATGGACCTGCCCCCGCAGCTCTCCTTCGGCCCTCTATGTGGCCGCCTTTGCGCTGGGCTTC 60
         ||||||||||||||||||||  |||||||||| ||| | ||||  ||||| ||||||||||
Db   129 ATGGACCTGCCCCCGCAGCTCTTCTTCGGCCCTCTATGTGGCCGCCTTCGCGCTGGGCTTC 188

Qy    61 CCGCTCAACCTCCTGGCCATCGAGGCGGCACGCGCGCCACCGGCTCGGTCGTCCACCCCCT 120
         |||||||||||||||| |||||  ||||| | ||| ||||| |||||||||||||||||||
Db   189 CCGCTCAACATCCTGGCCATCGGCGCCGTGTCCAGCGCGCCGTCCCAGCCCGGCCTCACCCCC 248

Qy   121 AGCCTGGTCTACGCCCTGAACCTGGCTGCTC 180
         ||||||  |||||| |||| |||| |||| |||
Db   249 AGCCTCGTTCGTTCACGCCCTCACCTGGGCTTCTGTGACCTCCGTCGTCGACACCTGCCCC 308
```

```
Query Match                    32.7%;   Score 295.2;  DB 14;  Length 609;
Best Local Similarity          90.5%;   Pred. No. 6.3e-42;
Matches   315;  Conservative     0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy    1  ATGGACCTGCCCCGCAGCTCTCCTTCGGCCTCTCTATGTGGCGCCTTGCCGGTGGCTTC   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  262  ATGGACCTGCCCCGCAGCTCTCCTTCGGCCTCTCTATGTGGCGCCTTGCCGGTGGCTTC  321

Qy   61  CCGCTCAACGTCCTGCCATCCGAGGCGCCACGGCCCACGCCCGGCTCCGGTCTCACCCCT  120
         |||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db  322  CCGCTCAACGTCCTGCCATCCGAGGCGCACGGCGCACGCCCGGCTCCGGTCTCACCCCT  381

Qy  121  AGCCTGGTCTACGCCCTGAACCTGGCTGCTGTCTGCGACCAGTCTCTGCCC          180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  382  AGCCTGGTCTACGCCCTGAACCTGGCTGCTGTCTGCGACCAGTCTCTGCCC          441

Qy  181  CTGAAGGCGGTGGAGGCGCTAGCCTCCGGGGCCTCGTGTGCCC                  240
         |||||||||||  ||||||||||  |||||||  ||  || |||| |||
Db  442  CTGAAGGCGGGGGAGGCGCTAAGCTCCGGGACCTGCCGACTCGCTGTGCGCCC        501

Qy  241  GTCTTCGCGGTGGCCCACTTCTTCCCACTCTATGCCGGGGGGGGCTTCCTGGCCCCCTG  300
         ||||||||||||||  |||| ||||||||||||| ||||| ||| ||||||||  ||| |
Db  502  GTCTTCGCGGTGGGCCAATCCTTCCCACTCTATTCCCGGGAAGTTTCCTGCCTCCCCG  561

Qy  301  AGTGCAGGGCCGCTACCTGGGGAGCAGCCTTCCCCTTGGGCTACCAAGCC          348
         ||  || |||| ||||||||| ||| ||   |||| ||||||||||| ||
Db  562  AGCGAAGCCCGACCCCTGGTAGCCCAGCCTGCCCTAGGGCTACCAACCC          609
```

```
RESULT 7
CE038975/c
LOCUS       CE038975              234 bp    DNA     linear    GSS 24-SEP-2003
DEFINITION  tigr-gss-dog-1700034974378 Dog Library Canis familiaris genomic,
            genomic survey sequence.
ACCESSION   CE038975
VERSION     CE038975.1  GI:35069198
KEYWORDS    GSS.
SOURCE      Canis familiaris (dog)
ORGANISM    Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1  (bases 1 to 234)
AUTHORS     Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
            Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
            Venter,J.C.
TITLE       The dog genome: survey sequencing and comparative analysis
JOURNAL     Science 301 (5641), 1898-1903 (2003)
MEDLINE     22875432
PUBMED      14512627
COMMENT     Contact: Kirkness EF
            The Institute for Genomic Research
            Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
            Rockville, MD 20850, USA
            Tel: 301-838-0200
            Fax: 301-838-0208
            Email: ekirknes@tigr.org
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1..234
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /strain="Standard Poodle"
                     /db_xref="taxon:9615"
                     /clone_lib="Dog Library"
                     /note="Site 1: BstXI; Libraries were prepared from
                     peripheral blood"
ORIGIN

Query Match                    21.0%;   Score 189.8;  DB 29;  Length 234;
Best Local Similarity          88.4%;   Pred. No. 1.6e-23;
Matches   206;  Conservative     0;  Mismatches   27;  Indels    0;  Gaps    0;
```

```
Qy  181  CTGAAGGCGGTGGAGGCGCTAGCCTCCGGGGCCTCGTGTGCCC                  240
         |||||||||||||||||||||   ||   ||||||||||||||||
Db  309  CTGAAGGCGGTGGAGGCGCTGGCCCTGGCCCCGGGGCCTCGTGTGCCC            368

Qy  241  GTCTTCGCGGTGGCCCACTTCTTCCCACTCTATGCCGGGGGGGGCTTCCTGGCCCCCTG  300
         ||||| ||| ||||||||||| | ||| ||   |||||||||||||  |||| ||  | |
Db  369  GCCTTTGCCCTGGCCCACTTCGCGGCCCCTACGGCGGGGGGGGCTTCCTGGTGCCCCTG  428

Qy  301  AGTGCAGGGCCGCTACCTGGGAGCAGCGTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCG  360
         ||  | ||| |||||||||||  |||||  |||||||||||||||||||  |||||||||
Db  429  AGCGTGGCCGCTACCTGGGGGCGCCCCCTTGGGCTACCAAGCCGCCCGCAGGCCG       488

Qy  361  TGCTATTCCTGGGGGGTGTGCGCGGCCATCTGGG                          394
         |||||| ||||||||| |||| ||||||| | |||
Db  489  CGCTACTCCTGGGGCCGTGTGTGCGGCCCATATGGG                        522
```

```
RESULT 6
CA776599
LOCUS       CA776599              609 bp    mRNA     linear    EST 03-DEC-2002
DEFINITION  ip08c01.y1 HR85 islet Homo sapiens cDNA clone IMAGE:6216192 5'
            similar to SW:GP40_HUMAN O14842 PUTATIVE G PROTEIN-COUPLED RECEPTOR
            GPR40. ;, mRNA sequence.
ACCESSION   CA776599
VERSION     CA776599.1  GI:26014474
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 609)
AUTHORS     Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J.,
            Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R.,
            Williams,T., Jackson,Y. and Bowers,Y.
TITLE       Endocrine Pancreas Consortium
JOURNAL     Unpublished (2000)
COMMENT     Other ESTs: ip08c01.x1
            Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Dr. Hiroshi Inoue
            (hinoue@im.wustl.edu)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 426.
FEATURES             Location/Qualifiers
     source          1..609
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6216192"
                     /tissue_type="Purified pancreatic islet"
                     /lab_host="DH10B"
                     /clone_lib="HR85 islet"
                     /note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
                     NotI; Site_2: XhoI; cDNA made by oligo-dT priming.
                     Size-selected on agarose gel. Average insert size ~1kb. 5'
                     XhoI site was destroyed after directional cloning.
                     Amplified once. Contact information: Hiroshi Inoue, MD,
                     Metabolism Div. (Alan Permutt Lab), Washington University
                     School of Medicine, Box 8127, 660 South Euclid Ave., St.
                     Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
                     314-362-1916, Fax: 314-747-2692."
ORIGIN
```

```
Matches 252;  Conservative   3;  Mismatches 159;  Indels   6;  Gaps   2;

Qy      20  TCTCCTTCGGCTCTATGTGGCCGGCCGTTGCGCTGGGCTTCCCGCTCAACGTCCTGGCCA   79
            ||||| | ||||   |||| |||  ||||||| |||||||||  |||| ||||||||||
Db     522  TCGGTCTGGGCGGCGTCTTCACCTTCCTCCTGGGCTGCCCGCCAACCTGGTCTGGTCC   581

Qy      80  TCCGAGGGCGGCACGGCCCAGCCCG---GCTCCGTCTCACCCTAGCCTGGTCTACGCCC   136
            ||||| || |||||||||||||||   || | | ||  |||| |||  ||||||||::|
Db     582  TCTTCGTTACCTGGCACAAGGCTGCAAGGCMGGCCACGCCCAACGTGTGTACGCMC   641

Qy     137  TGAACCTGGCTGTCTCCGACCTGTGCTGACAGTCTCTTGCCCCTGAAGGCGGTGGAGG   196
            |||||  |||  |||  |||| ||| |||||| |||| |||||| |||| |||| |||
Db     642  TCAACCTGTGCTGGCCAACTGGCCCTGTGTGGCCTGTGCCCTCGCTCAAGGCCCTGGAGA   701

Qy     197  CGCTAGCCTCCGGGCCTGGCCCTGCCGCCGGCCTCCTGCGCCCTGAGTGCAGGCCGCTACC   256
            ||| |  ||||||| | |||||||| | ||||||  |||| ||||||||||||||||||
Db     702  CGCT---GCTGCAGGACTGAGGCTGCCGCCAGGCGTCTGCCCGCCGCCTACAGCTTCTTCC   758

Qy     257  ACTTCTTCCCACTCTATGCCCGGCGGGGGCTTCCTGGCCCTGAGTGCCGGCCCGCTACC   316
            ||| |||||  |||  |||| |||  ||||  ||||  | ||||||| |||||||||||||
Db     759  TCTACTCCTGCTTCACGGCAGCTGCCTCCTTCCTCACGGCGCGCGGCCGGGGGCCGCTACC   818

Qy     317  TGGGAGCAGGCTTCCCCTTGGGCTACCAAGGCCTTCCGGAGGCGCCGTGCTATTCCTGGGGGG   376
            |||  ||||  ||  ||||||||| | ||||||  ||||  |||  | |  |||  |||||
Db     819  TCAGCATCGCTTCCCCATCGCTTCCATCCGCTACAAGCGCTACACCGGCGCGCTCGTCTTGG   878

Qy     377  TGTGCGGCGGCCATCTGGGCCCTCAGCGTGTCGTGTCACCTGGGTCTGTCTTTGGGTTGGAGG   436
            |  | |||||||||  |||| |||  ||| | ||| || |||| |||| | |||| ||||
Db     879  TCAGGCGCGGCTGTGGCTCAGCGTGGCTGTGCACCTGGCGCGGGCTGAGCGTGGGGCTGTGGGGARG   938
```

RESULT 9
B2160914

LOCUS   B2160914   774 bp   DNA   linear   GSS 11-OCT-2002
DEFINITION CH230-339G19.TJ CHORI-230 Segment 2 Rattus norvegicus genomic clone
            CH230-339G19, genomic survey sequence.
ACCESSION   B2160914
VERSION   B2160914.1 GI:23801870
KEYWORDS   GSS.
SOURCE   Rattus norvegicus (Norway rat)
  ORGANISM Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 774)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
            Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,
            Riggs,F., de Jong,P. and Fraser,C.M.
  TITLE   Rat BAC End Sequences from Library CHORI-230 MboI segment
  JOURNAL   Unpublished (1999)
COMMENT   Other GSSs: CH230-339G19.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 339   row: G column: 19
            Seq primer: SP6
            Class: BAC ends.
FEATURES   Location/Qualifiers
     source   1..774
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"

```
Qy     470  CCCTGGGCGATCAACACCGGTCAACGGCTCTCCGGTCTGCCTGGGAGGCCTGGGACCCG   529
            |||||||| |||  || ||||||||||  ||| |||||||| | ||||||||||||||
Db     234  CCCTGGGCAGCATCAACACCGGTCAACGGCTCTCCAGTCTGCCTGGGAGGCCTGGGACCCTG   175

Qy     530  CCCTGCGCCCGGCCCGCCGCCTTCAGCCTCTTCTCTCTCTTTTTTGCCCTGGCCA   589
            |||| | ||| ||||  |||||||| || || |||| || ||  |||||||||||
Db     174  CCTCAGCGGCGGCCCCGCCCGCCTCACCACTCTCGCTCTCCGCTCTTCCTGCCCCTGACCA   115

Qy     590  TCACAGCCTTCTGCTACGTGGCGGCTGCCTCCGGGCACTGCCCCGCTCCGCCGTGACGCACA   649
            ||||| | |||||||||| |||| || ||||||||||||| | ||| || |||| ||||||
Db     114  TCACGGCCTTCTGCTACGTGGGCTGCCTGCCGGGCACTGCCCCGCTCAGGCCTGAGCGCACA   55

Qy     650  GGGCGGAGCTGCGGGCCGCCGGGTGGGGCCGGCGGCCCCTCCTCACGCTGCTG   702
            |||  ||||  ||||||||| | |||||||| | |||||| |||| |||||||| 2
Db      54  GACGGAAGCTAAGGCAGCAGAACCTGGGGCGGGGGGGGCGGGGCCCCTGCTTCACGCTGCTG

RESULT 8
CNS041EX

LOCUS       CNS041EX                    961 bp     DNA     linear     GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            074E14 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL270114
VERSION     AL270114.1  GI:7992019
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat.Genet. 25 (2), 235-238 (2000)
   MEDLINE  20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
   MEDLINE  20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 961)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1..961
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="074E14"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG074BC07SP1-end :
                     PUC-Ori"
ORIGIN

Query Match          14.7%;  Score 132.4;  DB 29;  Length 961;
Best Local Similarity 60.0%;  Pred. No. 2.9e-13;
```

Left column

```
                    /clone="CH230-339G19"
                    /sex="Female"
                    /cell_type="Brain"
                    /clone_lib="CHORI-230 Segment 2"
                    /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                    CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                    Pieter de Jong"

ORIGIN
Query Match          14.4%;     Score 130;  DB 28;   Length 774;
Best Local Similarity 74.4%;    Pred. No. 7.1e-13;
Matches  177;  Conservative   0;  Mismatches  60;  Indels    1;  Gaps    1;

Qy     1 ATGGACCTGCCCCCGCAGCTCTCCTTCGGCCTCTCATGTGGCCGCCTTTGGCGCTGGGCTTC    60
         ||||||||||||||||||||||||||||||||||||||| |||||| || ||||||||||
Db   526 ATGGACCTGCCCCCACAGCTCTCCTTCGCTCTCATGTATCAGCCTTTGCACTAGGCTTT   585

Qy    61 CCGCTCAACGTCCTGGCCATCCGAGGCGCGACGACGCCCACGCCTCCGTCTCACCCCT   120
         |||||||  ||||| ||||||||||||||||||||||||||| |||||||||
Db   586 CCATTGAACTTGTTAGCCATCCGAGGTGCAGTGTCCCACGCGAAACTGGGACTCACCCCC   645

Qy   121 AGCCTGGTCTACGCCCTGAACCTGGGCTGCTCC-GACCTGCTGCTGACAGTCTCTCTGCC   179
         ||||||||||||||||||||||||||||||||| ||||||||||| |||||| |||||
Db   646 AGCTTGGTCTACACTCTCCATTTGGGCTGTGTTCCTGACCTTCCTACTGGCCATCACCCTGCC   705

Qy   180 CCTGAAGGCGGTGAGGCGCTAGCCTCCGGGGCCTTGGCCTTCGCCGGCCTTGCTGTGC   237
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   706 CCTGAAGGCGTGGAGGCCGCCCTGGCTTTCTGGGTCTCGGGGCCCTGGCGCACTCCCCTTCTGC   763
```

RESULT 10
```
AQ119988
LOCUS       AQ119988              543 bp    DNA      linear    GSS 22-SEP-1998
DEFINITION  HS_3024_A1_H01_T7 CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3024 Col=1 Row=0, genomic survey
            sequence.
ACCESSION   AQ119988
VERSION     AQ119988.1  GI:3497154
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 543)
AUTHORS     Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
TITLE       Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE     99380589
PUBMED      10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3024  row: O  column: 1
            Class: BAC ends
            High quality sequence stop: 543.
FEATURES             Location/Qualifiers
     source          1..543
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=3024 Col=1 Row=0"
                     /sex="male"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
```

Right column

```
ORIGIN
Query Match          14.2%;     Score 128;  DB 28;   Length 543;
Best Local Similarity 98.5%;    Pred. No. 1.4e-12;
Matches  128;  Conservative   0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy     1 ATGGACCTGCCCCCGCAGCTCTCCTTCGGCCTCTCATGTGGCCGCCTTTGGCGCTGGGCTTC    60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   414 ATGGACCTGCCCCCGCAGCTCTCCTTCGGCCTCTCATGTGGCCGCCTTTGGCGCTGGGCTTC   473

Qy    61 CCGCTCAACGTCCTGGCCATCCGAGGCGCGACGACGCCCACGCCGCCGTCTCCGTCTCACCCCT   120
         ||||||||||||||||||||||||||||||||||||||||| ||||||| ||||
Db   474 CCGCTCAACGTCCTGGCCATCCGAGGCGCGACGGNCCACGCNCCGNCGTCCGTCTCACCCCT   533

Qy   121 AGCCTGGTCT   130
         ||||||||||
Db   534 AGCCTGGTCT   543
```

RESULT 11
```
BM544968
LOCUS       BM544968             1117 bp    mRNA     linear    EST 20-FEB-2002
DEFINITION  AGENCOURT_6497637 NIH_MGC_125 Homo sapiens cDNA clone IMAGE:5588752
            5', mRNA sequence.
ACCESSION   BM544968
VERSION     BM544968.1  GI:18776664
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1117)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/,
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Invitrogen
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12360  row: f  column: 17
            High quality sequence start: 26
            High quality sequence stop: 716.
FEATURES             Location/Qualifiers
     source          1..1117
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5588752"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_125"
                     /note="Organ: ovary (pool of 3); Vector: pCMV-SPORT6;
                     Site_1: EcoRV (destroyed); Site_2: NotI; RNA source pool
                     of three ovaries, from females ranging in age from 38 to
                     49 yo. Library is oligo-dT primed and directionally cloned
                     (EcoRV site is destroyed upon cloning). Average insert
                     size 2.1 kb, insert size range 1-3.5 kb. Library is
                     normalized and enriched for full-length clones and was
                     constructed by C. Gruber (Invitrogen). Research Genetics
                     tracking code 036."

ORIGIN
Query Match          13.0%;     Score 117.4;  DB 12;   Length 1117;
Best Local Similarity 52.1%;    Pred. No. 1.3e-10;
Matches  371;  Conservative   0;  Mismatches  317;  Indels   24;  Gaps    4;

Qy    18 GCTCTCCTTCGGCCTCTCATGTGGCCGCCTTTGGCGCTGGGCTTCCCGCTCAACGTCCTGGC    77
         ||||| ||||| || ||||| ||||||| ||||| |||||||||||
Db   125 GTTCGTCTTCTCGGTGTACCTTCACTTTCCTCGGTGTGGGCGCTCCCCCTCAACCTGCTGGC   184
```

found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: IRBK1 row: e column: 10
High quality sequence stop: 748.

FEATURES   Location/Qualifiers
source   1..779
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:6971747"
                /tissue_type="mixed"
                /lab_host="DH5A (T1 phage-resistant)"
                /clone_lib="NIH_MGC_195"
                /note="Vector: pDNR-Dual; Site_1: loxP-SalI; Site_2:
                loxP-HindIII; Clones from this library have been
                PCR-amplified using gene-specific primers to contain the
                complete open reading frame (based on known gene sequences
                available from NCBI's RefSeq). Template for PCR is cDNA
                derived from either pooled cytoplasmic polyA RNA from 30
                cells lines or pooled total RNA from 10 different tissues
                (from BD Biosciences/Clontech and Washington University).
                PCR products are directionally cloned into the loxP sites
                of the pDNR-Dual vector. Library constructed by Dr.
                Narayan Bhat, Earl Bere III and Hongling Liao (Gene
                Expression Laboratory, Research Technology Program, SAIC
                Frederick, NCI-Frederick, Frederick, MD 21702). For
                information on which gene each clone represents, please
                visit our anonymous ftp site at
                ftp://image.llnl.gov/image/rearrayed_plates/IRBK.preSV.dat
                a Note: this is a NIH_MGC Library."

ORIGIN

Query Match   11.8%; Score 106.6; DB 14; Length 779;
Best Local Similarity 51.0%; Pred. No. 9.2e-09;
Matches 343; Conservative   0; Mismatches 309; Indels 21; Gaps 3;

Qy   18 GCTCTCCTTCGGCCCTCTATGTGGCGCCGCCTTTGCGCTGGGGCTTCCCCGTCTCAACGTCTCGGC   77
          |||| | ||| || | ||| | | || ||||||| | || | ||| || | || ||| |||
Db   62 GTTCGTCTTCCGGTGTACCTTCCTCACTTTCCTGGTGCCCCCCCTCAACCTGCTGGC   121

Qy   78 CATCCGAGGCGCGACGGCCCACGCCCGGCTCCGTCCGTCTCACCCTGCTGGTCTACGCCCT   137
          | | || | | | || || | | ||| ||| || | |||| | | | || |
Db   122 CCTGGTGGTCTTCGTGGGCAAGCTGCAGCGCGCCGGTGGCCGTGGACGTGCTCCTGCT   181

Qy   138 GAACCTGGGCTGCCTCCGGACCTGACAGTCTCGCCGGCCCTGGGCCCCCTGAAGGGCGGTGGAGGC   197
          |||| || ||| ||| | | | || ||| | || | | ||| | | |||| || |
Db   182 CAACCTGACCGGCTCCGGGACCTGTGCTGTTCCTGCCCTTTTCCTGGGCTCGGTGCCCTCGGCCCA   241

Qy   198 GCTAGCCTGGGGACCTGGCCTTGCGGCCTCCGGGGCGGGGGCTTCTGGCCCGCCCCCTGAGTCTACCT   257
          | || | || || ||| || | ||| || | | || || | | ||| | || ||||
Db   242 AGCCAATGGCATGCACTGGCCCTTCATCCTCTGCGGCGGCCGCCGCGAGGTCTGGT   301

Qy   258 CTTCTTCCCACTCTATGCCCGGGGGCTTCTGGCCGCGCCCTGAGTGCAGGCGCTACCT   317
          ||||| ||| | || | | | | || || | || ||| | | | |||| |
Db   302 CTTCACCACTATCTCACCGCCCCTTCCTCCTGGCGGCGTGCGAGCATTGAACGCTTCCT   361

Qy   318 GGGAGCAGCCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCGTGCTATTCCTGGGGGGT   377
          || || || ||||| |||| ||||||||| |||||||| | | || ||| || |
Db   362 GGGGAGCAGCCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCGTGCTATTCCTGGGGGGT   421

Qy   378 GTGCGGCGGCCATCGGGCCCTGCGTCCGTGTCACCTGGGTCTGTTTGGGTTGGAGGC   437
          ||| |||| || | || || | ||||| | |||||| ||| ||||| ||| |||
Db   422 GTGCGGCGGCCATCGGGCCCTGCGTCCGTGTCACCTGGGTCTGTTTGGGTTGGAGGC   481

Qy   438 TCCAGGAGGTGGCTGGACCACAGCAACAACCTCCCTTGGGCATCAACACCGGTCAACGG   497
          |||||||||| ||| | ||| | || ||| ||| | || |||| | || ||||
Db   482 CTCAGGGG---------------ACATCTCCACAGCCAGGGCACCAATGGGGACCTG   523

Qy   498 CTCTCCGTTCGCTGGAGGCTGGACCCGGCCGCCGGCCCGTTCCAGCCT   557
          |||||| || | ||| | | || | || | || | |||| |
Db   524 CTACCTGGAGTTCCGGAAGGACCCAG--CTAGCCATCCTCCTCCTGCCGCCATCCTCCTGCGGCTGGAG-AT   580

Qy   558 CTCTCCCTGCTCTTTTTCTGCCCTTGGCCATCACAGCCTTCGCGGGGCTGCCT   617
          ||||||| || | | || ||| || ||| | | || || ||||||||||

RESULT 12
CD559556
LOCUS   CD559556
DEFINITION AGENCOURT 14496985 NIH_MGC_195 Homo sapiens cDNA clone
            IMAGE:6971747 5', mRNA sequence.
ACCESSION   CD559556
VERSION   CD559556.2 GI:38558994
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 779)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   On Jun 10, 2003 this sequence version replaced gi:31585624.
            Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Narayan Bhat
            cDNA Library Preparation: Bhat Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be

Qy   78 CATCCGAGGCGCGACGGCCCACGCCCGGCTCCGTCCGTCTCACCCTGCTGGTCTACGCCCT   137
Db   185 CCTGGTGGTCTTCGTGGGCAAGCTGCAGCGCGCCGGTGGCCGTGGACGTGCTCCTGCT   244

Qy   138 GAACCTGGGCTGCCTCCGGACCTGACAGTCTCGCCGGCCCTGGGCCCCCTGAAGGGCGGTGGAGGC   197
Db   245 CAACCTGACCGGCTCCGGGACCTGTGCTGTTCCTGCCCTTTTCCTGGGCTCGGTGCCCTCGGCCCA   304

Qy   198 GCTAGCCTGGGGACCTGGCCTTGCGGCCTCCGGGGCGGGGGCTTCTGGCCCGCCCCCTGAGTCTACCT   257
Db   305 AGCCAATGGCATGCACTGGCCCTTCATCCTCTGCGGCGGCCGCCGCGAGGTCTGGT   364

Qy   258 CTTCTTCCCACTCTATGCCCGGGGGCTTCTGGCCGCGCCCTGAGTGCAGGCGCTACCT   317
Db   365 CTTCACCACTATCTCACCGCCCCTTCCTCCTGGCGGCGTGCGAGCATTGAACGCTTCCT   424

Qy   318 GGGAGCAGCCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCGTGCTATTCCTGGGGGGT   377
Db   425 GAGTGTGGCCCACCCACTGTGGTACAAGACCCGGCCAGGGCCGGAGGTCGGT   484

Qy   378 GTGCGGCGGCCATCGGGCCCTGCGTCCGTGTCACCTGGGTCTGTTTGGGTTGGAGGC   437
Db   485 GAGTGTGGGCTGCTGGTGTGGCCTCGCTCCGCAGCGTGTCTACGTCTCATAGAATT   544

Qy   438 TCCAGGAGGTGGCTGGACCACAGCAACAACCTCCCTTGGGCATCAACACCGGTCAACGG   497
Db   545 -----------ACATCTCCACAGCCAGGGCACCAATGGGGACCTG   586

Qy   498 CTCTCCGTTCGCTGGAGGCTGGACCCGGCCGCCGGCCCGTTCCAGCCT   557
Db   587 CTACCTGGAGTTCCGGAAGGACCAG--CTAGCCATCCTCCTCCTGCCGCCATCCTCCTGCGGCTGGAG-AT   643

Qy   558 CTCTCCCTGCTCTTTTTCTGCCCTTGGCCATCACAGCCTTCGCGGGGCTGCCT   617
Db   644 GGCTGTGGTCCTCTTTGTGGTCCCGCTGATCATCACCAGCTACTGCTAC--AGCCGGCCT   700

Qy   618 CCGGGCACTGGCCGCCGCTGACGCCACAGCGGAAGCTGCGGCCGCCGCTGGGTGGC   677
Db   701 GGTGGATCCTCGGCAGAGGGGCAGCCACCGCGCGGCAGAGGAGGGTGCGGNGCTGTT   760

Qy   678 CGGCGGGGGCCCCTTCTCACGCGTGCTCTGGGTAGGACCCTACAAACGGCTCC   729
Db   761 GGCGGCCACGCTGCTCAACTTCCTTGTCTGCGCTTTTGGGCCCCTACACAACGTGTCC   812

Db   138 GAACCTGGGCTGTCGTCCGACCGTGCTGCTGACAGTCTCTGCCCCTGAAGGCGGTGGAGGC 197
          |||||| |||||||||||||||||||| ||||||||| ||||||||| |||||||||||
Qy   213 CAACCTGACCGGCCTCGGACCTGTCTCTGCTGCGTGTTCCGCATGGTGGAGGC 272

Qy   198 GCTAGCCTCCGGGCGCTCGGCCCTGTCGCGCTCGGTCTTTCGCGGTCGGGGCCA 257
          ||||||||||||||||| || |||| |||||||||||||||||||||||||||
Db   273 AGCCAATGGCATGCACTGGCCCTCATCCTCGCCCCACTCTCTGGATTCATCTT 332

Qy   258 CTTCTTCCCACTCTATGCCGGACGGGGGCTTCCTGGGCGCCCTGAGTGAGCCGCTACCT 317
          |||||||||||||||||||||||||||| || |||||||||||||||||||
Db   333 CTTCACCACCATCTATCCACGGCCCTTCCTGGCAGCTGTGAGCATTGAACGCTTCCT 392

Qy   318 GGGAGCAGCCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCGTGCTATTCCTGGGGGT 377
          ||||||||||| |||||||||||||||||||||||| ||||||||||||||
Db   393 GAGTGTGGCCCACCCACTGTGTACAAGAACCCCGGCCGAGGCTGGGGCCAGGCAGGTCTGT 452

Qy   378 GTGCGCGGCCATCTGGGCCCTCGTCTGTCCTGGTCTGCTGGGTCTTTGGGTTGGAGGC 437
          |||||| |||||||||||||||||||||||||
Db   453 GAGTGTGGCCCTGGCTGTCGTGTTGGCCTCTGCCTCACTCGACCGTGGTCTACGTCATAGAATT 512

Qy   438 TCCAGGAGGC 447
          ||| ||||
Db   513 CTCAGGGGAC 522

RESULT 14
BI541693
LOCUS   BI541693   472 bp   mRNA   linear   EST 30-AUG-2001
DEFINITION 455527 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BI541693
VERSION   BI541693.1 GI:15382805
KEYWORDS   EST.
SOURCE   Bos taurus (cow)
ORGANISM   Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 472)
AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,B., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C.,
            Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A.,
            Chitko-McKown,C.G., Pertea,G., Holt,I., Karamycheva,S., Liang,P.,
            Quackenbush,J. and Keele,J.W.
TITLE   Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
JOURNAL   Genome Res. 11 (4), 626-630 (2001)
MEDLINE   21180013
PUBMED   11282978
COMMENT   Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR Primers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 121 row: M column: 22
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES   Location/Qualifiers
source   1..472
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="MARC 1BOV"
                     /note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
                     Library made from pooled tissue from lymph node, ovary,

---

Db   581 GGCTCGTGGTCCTCTTCTTGTGTGTCCCGCTGATCATCACCAGTACTGCTACAGCCGCTGGT 640
          ||||| |||||||||||||||||||||||||||||||||||| |||| |||||||
Qy   618 CCGGGCACTGGCCGCCCGCTCCGGCTGAGACGACACAGAGCGGAAGCTGCGGGGCGCCTGGGTGGC 677
          |||||||||||||||||||||||||| |||| ||||||||||||| |||
Db   641 GTGGATCCTCCGGCAGAGGGGGCAGCCACCGGGCCACCGCGGCGGGAGGAGGTGGCGGGGCGTGTTGGC 700

Qy   678 CGGCGGGGCCTC 690
          ||| |||||
Db   701 GGCCACGCTGCTC 713

RESULT 13
CF147780
LOCUS   CF147780   696 bp   mRNA   linear   EST 25-JUL-2003
DEFINITION AGENCOURT 14740239 NIH_MGC_145 Homo sapiens cDNA clone
            IMAGE:6971953 5', mRNA sequence.
ACCESSION   CF147780
VERSION   CF147780.1 GI:33244048
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 696).
AUTHORS
TITLE   NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL   National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
COMMENT   Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: GPCR Consortium
            cDNA Library Preparation: GPCR Consortium
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: IRBI02 row: f column: 12
            High quality sequence stop: 680.
FEATURES   Location/Qualifiers
source   1..696
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6971953"
                     /tissue_type="mixed"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_145"
                     /note="Vector: pcDNA3.1; Site 1: varies by clone; Site 2:
                     varies by clone; pcDNA3.1 by the GPCR Consortium. Cloning sites vary by
                     clone and include the following: 5'-EcoRV-XmnI/XhoI-3',
                     5'-EcoRV-XmnI/NotI-3', EcoRV (TA cloned, non-directional).
                     For information about which gene each clones represents,
                     please visit our anonymous ftp site at
                     ftp://image.llnl.gov/image/rearrayed_plates/IRBI.preSV.dat
                     a Note: this is a NIH_MGC Library."

ORIGIN

Query Match   11.5%;   Score 103.6; DB 14; Length 696;
Best Local Similarity 52.6%; Pred. No. 3e-08;
Matches 226; Conservative 0; Mismatches 204; Indels 0; Gaps 0;

Qy   18 GTCTCCTTCCGGCCTCTATGTGGCCGCCCTTGCCGCTGGCTTCCCGGCTCAACGTCCTGGC 77
          || ||| ||| || |||||||||||||||||||||||||||||
Db   93 GTTCGTCTTTCTCGGTGTACCTTTCCCTGACTTTCCTGTGTGGGGCTCCCCAACCTTGCTGGC 152

Qy   78 CATCCGAGGCGCGACGGCCCCAGCGCCCGGCTCCCGTCTCACCCCTAGCCTGGTTCACGCCCT 137
          ||||||| || |||| ||||||||||||||||| |||| |||||||
Db   153 CCTGTGTGGTCTTCGTGGGCAAGCTGCAGCACCGGCCCGCGCGGGCCGTGGCCGTGGACGTGCTCCTGCT 212

ORIGIN   fat, hypothalamus, and pituitary."

```
Query Match           11.4%;    Score 102.8; DB 12;   Length 472;
Best Local Similarity 53.8%;    Pred. No. 3.7e-08;
Matches 212;  Conservative  0; Mismatches 182;  Indels  0; Gaps  0;

Qy    18 GCTCTCCTTCGGCCTCTCATGTGGCCGCCTTTGCGCTGGGCTTCCCGCTCAACGTTCCTGGC 77
          ||||||  |||| || ||  ||  | ||| ||| ||||||||| || |||||||||||||
Db    76 GCTCTTCTTCTCCGTGTACCTCTTTCACCTTCCTCGTGGGCTTCCCCCCTCAACCTGATGGC 135

Qy    78 CATCCGAGGCGGCGACGGGCCACGCCCGGCTCCGTCCTCACCCCTAGCCTGGTCTACGCCCT 137
          ||  |  |  |||  ||| || |||||||| |||| || | ||| || |||||||||||
Db    78 GGCCAGTGCCATGCACTGTCCCGTCCCCTTCGTCCCCCTTTCTGCCCCTTTCGTGGGCTTC 137

Qy   136 CCTGGTGATCTTCGTGGGCAAGCTGGCGGCCGCCGCCGGCGTGTGGACGTGCTCTTGCT 195
          || |||  |||| ||| || ||  || | || | ||||||
Db   136 CCTGGTGATCTTCGTGGGCAAGCTGGCGGCCGCCGCCGGCGTGTGGACGTGCTCTTGCT 195

Qy   138 GAACCTGGGCTGCTGCTCCGACCTGCTCTGCTGACAGTCTCTGCCCCTGAAGGCGGTGGAGGC 197
          ||||| || | | | |    | |||   || | |||||||||||||
Db   196 AAACCTCACCCTCTCGGATCTGGTCCTGTTGCTCGTTCCTGCCTGTTCCGCATGGTGGAGGC 255

Qy   198 GCTAGCCTCCGGGGGCTCGGCCTGGCCTTCGCCGGCCTCGTGTGCCCGTCTTCGCGGTGGCCCA 257
          ||  | |  | | |||| || || || || |||| || | |||||||||||
Db   256 GGCCAGTGCCATGCACTGTCCCGTCCCCTTCGTCCCCCTTTCTGCCCCTTTCTCCAGGTTCCTCTT 315

Qy   258 CTTCTTCCCACTCTCATGCGCGGGGGCTTCCTGGCCGCCCTCGAGTGAGTGCAGGCCGCTACCT 317
          ||| | ||| | ||||    |||||| |||    ||  | || ||| ||
Db   316 CTTCACCACCATCTATCTCACGTCCCCTCTTCCTGGCAGCCGTGGCAGCACAGAGCGCTTCCT 375

Qy   318 GGGAGCAGCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCGTGCTATTCCTGGGGGT 377
          |||  ||  |||||| |  | |||  || ||||    || | ||| |||||
Db   376 GAGCGGTGGCCTACCCGGCCTTGGTACAAGACTCCGGCCCGAGGCCAGGCCGCCTGGT 435

Qy   378 GTGCGCGGCCATCTGGGCCCTCGTCCTGTCCTGTGTCAC 411
          | |||||  ||||  |||||| ||| || | ||| |||
Db   436 CAGTGGGGGCTGCTGGCTGGCTCTGGCCGCGTGCTCAC 469
```

RESULT 15
CA777066
LOCUS   CA777066   489 bp   mRNA   linear   EST 03-DEC-2002
DEFINITION   ip02f08.y1 HR85 islet Homo sapiens cDNA clone IMAGE:6134918 5'
               similar to SW:GP40_HUMAN O14842 PUTATIVE G PROTEIN-COUPLED RECEPTOR
               GPR40.;, mRNA sequence.
ACCESSION   CA777066
VERSION   CA777066.1 GI:26014941
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 489)
AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
               Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
               Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
               Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J.,
               Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R.,
               Williams,T., Jackson,Y. and Bowers,Y.
TITLE   Endocrine Pancreas Consortium
JOURNAL   Unpublished (2000)
COMMENT   Other ESTs: ip02f08.x1
               Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
               Endocrine Pancreas Consortium
               Harvard University, Howard Hughes Medical Institute
               Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
               MA 02138
               Tel: 617-495-1812
               Fax: 617-495-8557
               Email: dmelton@iohp.harvard.edu
               Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
               Washington University Genome Sequencing Center For information on
               obtaining a clone please contact: Dr. Hiroshi Inoue
               (hinoue@im.wustl.edu)
               Seq primer: -40RP from Gibco
               High quality sequence stop: 64.

FEATURES   Location/Qualifiers
     source   1..489
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6134918"
                     /tissue_type="Purified pancreatic islet"
                     /lab_host="DH10B"
                     /clone_lib="HR85 islet"
                     /note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
                     NotI; Site_2: XhoI; cDNA made by oligo-dT priming.
                     Size-selected on agarose gel. Average insert size ~1kb. 5'
                     XhoI site was destroyed after directional cloning.
                     Amplified once. Contact information: Hiroshi Inoue, MD,
                     Metabolism Div. (Alan Permutt Lab), Washington University
                     School of Medicine, Box 8127, 660 South Euclid Ave., St.
                     Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
                     314-362-1916, Fax: 314-747-2692."

ORIGIN

```
Query Match           10.6%;    Score 95.8; DB 14;   Length 489;
Best Local Similarity 89.1%;    Pred. No. 6.3e-07;
Matches 115;  Conservative  0; Mismatches 12;  Indels  2; Gaps  1;

Qy     1 ATGGACCTGCCCCGCAGCTCTCCTTCGGCCTCTCATGTGGCCGCCTTTGCGCTGGGCTTC 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    94 ATGGACCTGCCCCGCAGCTCTCCTTCGGCCTCTCATGTGGCCGCCTTTGCGCTGGGCTTC 153

Qy    61 CCGCTCAACGTTCCTGGCCATCCGAGGCGGCGACGGGCCCACGCGCC--GGCTCCGTCTCACCC 118
          |||||||||||||||||||||||||||||||||||||  ||||||  ||||||||||||||||
Db   154 CCGCTCAACGTTCCTGGCCATCCGAGGCGGCGACGGGCCCGACGGGGCCACGCGCCGGGGTTCGTCTTCACCC 213

Qy   119 CTTAGCCTGG 127
          ||||| ||||
Db   214 TTAGCCTGG 222
```

Search completed: April 30, 2004, 16:32:52
Job time : 2953 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2004, 07:54:59 ; Search time 90 Seconds
                                       (without alignments)
                          5568.007 Million cell updates/sec

Title: US-10-202-687-1
Perfect score: 903
Sequence: 1 atggacctgcccccgcagct.........ggggcaagtcccagaagtaa 903

Scoring table: IDENTITY_NUC
              Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued Patents NA:*
      1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
      2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
      3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
      4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
      5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
      6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Query Length	DB	ID	Description
1	903	100.0	903	4	US-09-170-496D-249	Sequence 249, App
2	898.2	99.5	903	4	US-09-170-496D-271	Sequence 271, App
3	123.4	13.7	1841	2	US-08-820-521-1	Sequence 1, Appli
4	123.4	13.7	1841	3	US-09-248-715-1	Sequence 1, Appli
5	123.4	13.7	1841	3	US-09-248-715-1	Sequence 1, Appli
6	121.8	13.5	1041	4	US-09-170-496D-253	Sequence 253, App
7	118.6	13.1	1041	4	US-09-170-496D-273	Sequence 273, App
8	84.6	9.4	4895	3	US-09-053-866-1	Sequence 1, Appli
9	84.6	9.4	4895	4	US-09-479-130-1	Sequence 1, Appli
10	84.6	9.4	4895	4	US-09-472-130A-1	Sequence 1, Appli
11	80	8.9	1098	4	US-09-170-496D-117	Sequence 117, App
12	80	8.9	1098	4	US-09-170-496D-225	Sequence 225, App
13	80	8.9	1597	2	US-08-724-974A-1	Sequence 1, Appli
14	80	8.9	1697	4	US-09-364-425B-26	Sequence 26, Appl
15	78.2	8.7	1050	4	US-09-762-661A-1	Sequence 1, Appli
16	78.2	8.7	1329	4	US-09-152-060-25	Sequence 25, Appl
17	77.2	8.5	993	4	US-09-170-496D-257	Sequence 257, App
18	77.2	8.5	2260	2	US-08-788-750-1	Sequence 1, Appli
19	75.6	8.4	993	4	US-09-170-496D-275	Sequence 275, App
20	73.2	8.1	993	3	US-09-187-710-1	Sequence 1, Appli
21	70	7.8	1095	4	US-09-743-475-2	Sequence 2, Appli
22	70	7.8	1910	3	US-09-009-438-1	Sequence 1, Appli
23	70	7.8	1910	3	US-09-207-493-1	Sequence 1, Appli
24	70	7.8	4009	4	US-09-743-475-1	Sequence 1, Appli
25	66.8	7.4	1155	4	US-09-053-866-3	Sequence 3, Appli
26	66.8	7.4	1155	4	US-09-479-130-3	Sequence 3, Appli
27	66.8	7.4	1155	4	US-09-472-130A-3	Sequence 3, Appli

28	65.8	7.3	2911	3	US-08-981-825-5	Sequence 5, Appl
29	65.8	7.3	2911	3	US-09-480-784-5	Sequence 5, Appl
30	64.4	7.1	1092	3	US-09-077-675A-15	Sequence 15, Appl
31	64.4	7.1	1092	3	US-09-077-674-15	Sequence 15, Appl
32	64.4	7.1	3129	3	US-09-077-675A-14	Sequence 14, Appl
33	64.4	7.1	3129	4	US-09-077-674-14	Sequence 14, Appl
34	64.2	7.1	1926	4	US-09-249-585A-4	Sequence 4, Appl
35	64.2	7.1	1931	2	US-09-130-114-2	Sequence 2, Appl
36	63	7.0	1063	4	US-09-077-675A-1	Sequence 1, Appli
37	63	7.0	1063	4	US-09-077-674-1	Sequence 1, Appli
38	62.8	7.0	1089	4	US-09-170-496D-3	Sequence 3, Appl
39	62.8	7.0	1089	4	US-09-170-496D-165	Sequence 165, App
40	62.8	7.0	1365	4	US-09-016-434-1429	Sequence 1429, Ap
41	62.2	6.9	1174	3	US-09-034-985-1	Sequence 1, Appli
42	61.4	6.8	1029	4	US-09-077-675A-4	Sequence 4, Appl
43	61.4	6.8	1029	4	US-09-077-674-4	Sequence 4, Appl
44	61.2	6.8	1285	4	US-09-016-434-1366	Sequence 1366, Ap
45	59.6	6.6	1143	4	US-09-170-496D-61	Sequence 61, Appl

ALIGNMENTS

RESULT 1
US-09-170-496D-249
; Sequence 249, Application US/09170496D
; Patent No. 6555339
; GENERAL INFORMATION:
;   APPLICANT: Behan, Dominic P.
;   APPLICANT: Chalmers, Derek T.
;   APPLICANT: Liaw, Chen W.
;   TITLE OF INVENTION: No. 6555339-Endogenous, Constitutively Activated Human G Protein-C
;   TITLE OF INVENTION: Receptors
;   FILE REFERENCE: AREN-0040
;   CURRENT APPLICATION NUMBER: US/09/170,496D
;   CURRENT FILING DATE: 1998-10-13
;   NUMBER OF SEQ ID NOS: 294
;   SOFTWARE: Patentin version 3.1
;   SEQ ID NO 249
;   LENGTH: 903
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-170-496D-249

Query Match   100.0%; Score 903; DB 4; Length 903;
Best Local Similarity   100.0%; Pred. No. 2.3e-177;
Matches 903; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 ATGGACCTGCCCCCGCAGCTCTCCTTCGGCCTCATATGGGCCGGCCTTGGCCTGGGCTTC 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGGACCTGCCCCCGCAGCTCTCCTTCGGCCTCATATGGGCCGGCCTTGGCCTGGGCTTC 60

Qy   61 CCGCTCAACGTCCTGCCCATCGAGGCGCGACGGCCCACGCCGGCTCCGTCGTCCACCCCT 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CCGCTCAACGTCCTGCCCATCGAGGCGCGACGGCCCACGCCGGCTCCGTCGTCCACCCCT 120

Qy   121 AGCCTGGTCACGCCCTGAACCTGGACTCCGACAGTCTCTGCCC 180
            ||||||||||||||||||||||||||||||||||||||||||||
Db   121 AGCCTGGTCACGCCCTGAACCTGGACTCCGACAGTCTCTGCCC 180

Qy   181 CTGGAAGGCGGTGGAGGCGCTAGCCTCCGGGGCCTGCCCC 240
            ||||||||||||||||||||||||||||||||||||||||
Db   181 CTGGAAGGCGGTGGAGGCGCTAGCCTCCGGGGCCTGCCCC 240

Qy   241 GTCTTCGGCGGCCACACTTCTTCCCACTCTATGCCGCGGCCTTCCTGCCGCCGCCTG 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GTCTTCGGCGGCCACACTTCTTCCCACTCTATGCCGCGGCCTTCCTGCCGCCGCCTG 300

Qy   301 AGTGCAGGCGCGCTACCTGGGAGCAGCCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCG 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 AGTGCAGGCGCGCTACCTGGGAGCAGCCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCG 360

Qy   361 TGCTATTCCTGGGGGTGTGCGCGGCCATCTGGGCCCTCGTGTCACCTGGGTCTG 420

```
Qy  121 AGCCTGGTCTACGCCCTGAACCTGGCTGCTGCTCCGACCTGCTGCTGACAGTCTCTGCCC 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 AGCCTGGTCTACGCCCTGAACCTGGCTGCTGCTCCGACCTGCTGCTGACAGTCTCTGCCC 180

Qy  181 CTGAAGGCGGTGGAGGCGCTAGCCTCCGGGGCCTGGCTCTGCCGGGCCTGTGTCCCC 240
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 CTGAAGGCGGTGGAGGCGCTAGCCTCCGGGGCCTGGCTCTGCCGGGCCTGTGTCCCC 240

Qy  241 GTCTTCGCGGTGGCCCACTTCTTCCCACTCTATGCGCGGGGCTTCTTGGCCGCCCCTG 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 GTCTTCGCGGTGGCCCACTTCTTCCCACTCTATGCGCGGGGCTTCTTGGCCGCCCCTG 300

Qy  301 AGTGCAGGCCGCTACCTGGGAGCAGCCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCG 360
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 AGTGCAGGCCGCTACCTGGGAGCAGCCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCG 360

Qy  361 TGCTATTCCTGGGGGTGTGCGGCGACCTGCTCCTGTGTCACCTGGGTCTG 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 TGCTATTCCTGGGGGTGTGCGGCGACCTGCTCCTGTGTCACCTGGGTCTG 420

Qy  421 GTCTTTGGGTTGGAGGCTCGCTGCCTGGAGGCGCTCCTGCGCCTGCGCATC 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 GTCTTTGGGTTGGAGGCTCGCTGCCTGGAGGCGCTCCTGCGCCTGCGCATC 480

Qy  481 AACACACCGGTCAACGGCTCTCCGGCGCTCTCTTTTTTCTGCCCATCACAGCCTTC 540
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 AACACACCGGTCAACGGCTCTCCGGCGCTCTCTTTTTTCTGCCCATCACAGCCTTC 540

Qy  541 CCGGCCCGCTTCAGCCTGCTGCCCGGCACTGGCCCGCTCCGCAGCGGCGAAGCTG 600
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 CCGGCCCGCTTCAGCCTGCTGCCCGGCACTGGCCCGCTCCGCAGCGGCGAAGCTG 600

Qy  601 TGCTACGTGGGGCTGCCTCCGGGGCCTCCTCCAGCGGCCCTCGCGTAGGAGCCCTAC 660
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 TGCTACGTGGGGCTGCCTCCGGGGCCTCCTCCAGCGGCCCTCGCGTAGGAGCCCTAC 660

Qy  661 CGGGGCCGCCTGGGTGCCGGCGGGGCCCGCGCTCCTCCGCGGAGGCTCCTGGCGAAGCTG 720
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 CGGGGCAAATGGGTGCCGGCGGGGCCCGCGCTCCTCCGCGGAGGCTCCTGGCGAAGCTG 720

Qy  721 AACGCCTCCAACGTGGCCAGCTTGCCAGCTTCCTGTACCCCAATCTAGGAGGCGAAGCTG 780
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 AACGCCTCCAACGTGGCCAGCTTGCCAGCTTCCTGTACCCCAATCTAGGAGGCGAAGCTG 780

Qy  781 GGGCTCATCACGGGTGCTGCTGGAGTGGTGCTTAATCCGCTGGTGTGACCGGTTACTTGGGA 840
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 GGGCTCATCACGGGTGCTGCTGGAGTGGTGCTTAATCCGCTGGTGTGACCGGTTACTTGGGA 840

Qy  841 AGGGGTCCTGCCTGAAGACAGTGTGCGCAAGAACGCAAGGGGCAAGTCCCAGAAG 900
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  841 AGGGGTCCTGCCTGAAGACAGTGTGCGCAAGAACGCAAGGGGCAAGTCCCAGAAG 900

Qy  901 TAA 903
        |||
Db  901 TAA 903
```

RESULT 3
US-08-820-521-1
; Sequence 1, Application US/08820521
; Patent No. 5942416
; GENERAL INFORMATION:
; APPLICANT: Bergsma, Derk
; APPLICANT: Ganesh, Sathe
; APPLICANT: Fuetterer, Wendy
; APPLICANT: Mao, Joyce
; TITLE OF INVENTION: cDNA CLONE HNFDY20 THAT ENCODES
; TITLE OF INVENTION: A NOVEL HUMAN 7-TRANSMEMBRANE RECEPTOR
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia

```
        ||||||||||||||||||||||||||||||||||||||||||
Db  361 TGCTATTCCTGGGGGTGTGCGGCGACCTGCTCCTGTGTCACCTGGGTCTG 420

Qy  421 GTCTTTGGGTTGGAGGCTCGCTGCCTGGAGGCGCTCCTGCGCCTGCGCATC 480
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 GTCTTTGGGTTGGAGGCTCGCTGCCTGGAGGCGCTCCTGCGCCTGCGCATC 480

Qy  481 AACACACCGGTCAACGGCTCTCCGGCGCTCTTTTTTCTGCCCATCACAGCCTTC 540
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 AACACACCGGTCAACGGCTCTCCGGCGCTCTTTTTTCTGCCCATCACAGCCTTC 540

Qy  541 CCGGCCCGCTTCAGCCTGCTGCCCGGCACTGGCCCGCTCCGCAGCGGCGAAGCTG 600
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 CCGGCCCGCTTCAGCCTGCTGCCCGGCACTGGCCCGCTCCGCAGCGGCGAAGCTG 600

Qy  601 TGCTACGTGGGGCTGCCTCCGGGGCCTCCTCCAGCGGCCCTCGCGTAGGAGCCCTAC 660
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 TGCTACGTGGGGCTGCCTCCGGGGCCTCCTCCAGCGGCCCTCGCGTAGGAGCCCTAC 660

Qy  661 CGGGGCCGCCTGGGTGCCGGCGGGGCCCGCGCTCCTCCGCGGAGGCTCCTGGCGAAGCTG 720
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 CGGGGCCGCCTGGGTGCCGGCGGGGCCCGCGCTCCTCCGCGGAGGCTCCTGGCGAAGCTG 720

Qy  721 AACGCCCTCCAACGTGGCCAGCTTCCTGTACCCCAATCTAGGAGGCTCCTGGCGAAGCTG 780
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 AACGCCCTCCAACGTGGCCAGCTTCCTGTACCCCAATCTAGGAGGCTCCTGGCGAAGCTG 780

Qy  781 GGGCTCATCACGGGTGCCTGGAGTGTGGTGCTTAATCCGCTGGTGACCGGTTACTTGGGA 840
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 GGGCTCATCACGGGTGCCTGGAGTGTGGTGCTTAATCCGCTGGTGACCGGTTACTTGGGA 840

Qy  841 AGGGGTCCTGCCTGAAGACAGTGTGCGCAAGAACGCAAGGGGCAAGTCCCAGAAG 900
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  841 AGGGGTCCTGCCTGAAGACAGTGTGCGCAAGAACGCAAGGGGCAAGTCCCAGAAG 900

Qy  901 TAA 903
        |||
Db  901 TAA 903
```

RESULT 2
US-09-170-496D-271
; Sequence 271, Application US/09170496D
; Patent No. 6555339
; GENERAL INFORMATION:
; APPLICANT: Behan, Dominic P.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: No. 6555339-Endogenous, Constitutively Activated Human G Protein
; TITLE OF INVENTION: Receptors
; FILE REFERENCE: AREN-0040
; CURRENT APPLICATION NUMBER: US/09/170,496D
; CURRENT FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 271
; LENGTH: 903
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-170-496D-271

Query Match   99.5%; Score 898.2; DB 4; Length 903;
Best Local Similarity 99.7%; Pred. No. 2.3e-176;
Matches 900; Conservative 0; Mismatches 0; Indels 3; Gaps 0;

```
Qy    1 ATGGACCTGCCCCCCGCAGCTCTCCTTCGGCCTCTATGTGGCCCGCCTTTGGCTGGCTTC 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGGACCTGCCCCCCGCAGCTCTCCTTCGGCCTCTATGTGGCCCGCCTTTGGCTGGCTTC 60

Qy   61 CCGGCTCAACGTGCTGGCCATCCGAGGGCGCGACGCGCCCAGCCCCGCTCGTCTCACCCCT 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CCGGCTCAACGTGCTGGCCATCCGAGGGCGCGACGCGCCCAGCCCCGCTCGTCTCACCCCT 120
```

; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/820,521
; FILING DATE: 19-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Han, William T
; REGISTRATION NUMBER: 34,344
; REFERENCE/DOCKET NUMBER: GH50011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5219
; TELEFAX: 610-270-4026
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1841 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-820-521-1

Query Match   13.7%;   Score 123.4; DB 2; Length 1841;
Best Local Similarity 52.3%;   Pred No. 5.2e-17;
Matches 383; Conservative 0; Mismatches 326; Indels 24; Gaps 4;

Qy   18 GCTCCTCCTTCGGCCTCTATGTGGCCGCCTTTGCGCTGGGCTTCCCGCTCAACGTCCTGGC 77

Db   334 GTTCGTCTTCTCGGTGTACCTTTCTCACTTTCCTGTGGGGCTCCCCTCAACCTGCTGGC 393

Qy   78 CATCCGAGGCGGACGACGGCCCACGCGCCGGCTCCGTCTCACCCCTAGCCTGGTCTACGCCCT 137

Db   394 CCTGGTGGTCTTCGTGTGGGCAAGCTGCAGCGTGCGTGGCCGTGGACGTGCTCCTGCT 453

Qy   138 GAACCTGGGCGTGCTGTCCGACCTGGTCTTCTCTGCCCCTGAAGCGGCGGTGGAGGC 197

Db   454 CAACCTGACCGCCTGGAACCTGGTCCTGCTGCGTGTTCCTGCCCTTTCCGGATGGTGGGAGGC 513

Qy   198 GCTAGCCTCCGGGGGCCTGGCCTTGGCCTCGCTGTGCCCCGTCTTGCGGTGGCCCA 257

Db   514 AGCCAATGGCATGCACTGGCCCCTGGCCCCTGCCCATCCTCTGCCACTCTCTGGATTCATCTT 573

Qy   258 CTTCTTCCCACTCTTCATGCCCGGCCGGGGGGCTTCCTGGCCGCCCTGAGTGACTGGCCGCTACCT 317

Db   574 CTTCACCACCACCATCTATCTCACCGCCCTTCCTGCAGCTGTGGAGCATTGAACGCTTCCT 633

Qy   318 GGGGAGCAGCCTTCCCCTTGGCTACCAAGGCTTCGGAGGCCGTGCTATTCCTGGGGGGT 377

Db   634 GAGTGTGGCCCACCCCCTGTGGTACAAGAACCCGGCCTGGGGCAGGCCAGTCTCGT 693

Qy   378 GTGCGCGGCCATCTGGGCCCCTGTCCTGTGTCACCTGGGTCTTTGGGTTGGAGGC 437

Db   694 GAGTGTGGCCTGCTGTGGCCTGTTGGCGTGGGCTCGCAGCGTGGTCTACGCGTGCCTGCTCATAGAATT 753

Qy   438 TCCAGGAGGCTGGCTGGACCACAGCAACCACCTCCCTGGGCATCAACACACCGGTCAACCGG 497

Db   754 CTCAGGGG---------ACATCTCCCACAGCAGCGCCAGGGCCCACCAGGGGACCTG 795

Qy   498 CTCTCCGGTGCTGCCCTGGAGGCCCGCTCTGGGACCCGGCTTCCCGCCCGCCTTCAGCCT 557

Db   796 CTACCTGGAGTTCCGGAAGGACCAG--CTAGCCCATCCTCCTGCCCGTGCCCGTGGGAG-AT 852

Qy   558 CTCTTCCCTGCTCTTTTTTCTGCCCTTGGCCATCACAGCGCCCTTCTGCTACGCGTGGGCTGCCT 617

APPLICATION NUMBER: US/09/248,715
FILING DATE: 09-Feb-1999
CLASSIFICATION: UNKNOWN
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/820,521
FILING DATE: 19-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Prestia, Paul F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GP-50011-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0700
TELEX: 846169
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1841 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-248-715-1

Query Match   13.7%;   Score 123.4;   DB 3;   Length 1841;
Best Local Similarity 52.3%;   Pred. No. 5.2e-17;
Matches 383;   Conservative   0;   Mismatches 326;   Indels 24;   Gaps   4;

```
Qy   18 GCTCTCCTTCGGCCTCTATGTGGCCGGCTTCCCGGCTCAACGTCCTGGC 77
           ||  | || |||  ||| | |||||||    ||||    ||||||||||
Db  334 GTTCGTCTTCTCGGTGACCTTCTCACTTTCCTGTGTGGGGCTCCCCTCAACCTGCTGGC 393

Qy   78 CATCCGAGGCGGACGGCCCACGGCCCGTCGTCTCAACCCCTCAGCCTGTCTACGCCCCT 137
           ||||  |||  |  ||||||||||||||||
Db  394 CCTGGTGGTCTTCGTGGCAAGCTGCAGAGTCTTCTGTGTCCCTGCCTTTCTGCCTCCTGCT 453

Qy  138 GAACCTGGGCTGCTCCGACCTGCTGCGCGGGGGGCTTCCTGGCCGCCGCTGCTTCCTGCTACT 197
           |||||||||||||||||||
Db  454 CAACCTGACCGACCTGCTCCTGTGTCCGTGCTTCTTGCCTTTCCGCATGGTGGAGGC 513

Qy  198 GCTAGCCTCCGGGGCCTGGACCTGCTCGGTCTCGGTCTCTTTGGGTTGGAGGC 257
           ||||| | ||| ||||||||| |||
Db  514 AGCCAATGGCATGCACTGGCCCCTCATCCTCTGCGCCCACTCTCTGATTCATCTT 573

Qy  258 CTTCTTCCCACTCTATGCCGCCCGGCTTCCTGGCCGCCGCTGCTACCT 317
           ||  || ||||||| | ||| | ||||||||||||||
Db  574 CTTCACCACCATCTATCTCACCGCGCCCTTCCTGGCAGCTGTGAGCATTGAACGCTTCCT 633

Qy  318 GGGAGCAGCTTCCCTTGGGCTACCAAGCCTTCGGAGGCCGTGCTATTCCTGGGGGT 377
           |||| |||| |||||||  |||||
Db  634 GAGTGTGGCCCACCCCCCCTGTGGCCCCGTCGTCCTGCGGCTGCGGCCTGGAG-AT 693

Qy  378 GTGCGGCGGCCATCTGGGCGCCTCGTCACCTGGTCTGGTTGGGTTGGAGGC 437
           |||||||||||||||||| || | |||||||||
Db  694 GAGTGTGGTCTTGTGGCCTGTGTGGCCTCACTGCAGCGTGGCTGGTTACGTCATAGAATT 753

Qy  438 TCCAGGAGGCTGGCTGGACCAGCAGCACCACACCTCCCTGGGCATCAACACACCGGTCAACGG 497
           || | ||
Db  754 CTCAGGGG-----------ACATCTCCACAGCCAGGCACCAATGGGACCTG 795

Qy  498 CTCTCCGGTCTGCCTGGAGGCCTGGACCCGCGCCTGCGCCGCCTTCAGCCT 557
           ||||    || |  || || || |
Db  796 CTACCTGGAGTTCCGGAAGGACCAG--CTAGCCATCATCCTCCTGCGCCTGGAG-AT 852

Qy  558 CTCTCCTGCTCTTTTTCGCCCTTGGCCATCACAGCCTTCTGCTACGTGGGCTGCCT 617
           ||||| ||||||| |
Db  853 GGCTGTGGTCCTCTTTTGTGGTCCCGCTGATCATCACCAGCTACTGCTAC---AGCCGCCCT 909

Qy  618 CCGGGCACTGGCCTCGGCGCCCGCTGACGCGAAGCTGCGGGCCCGCTGGGTGGC 677
           || | ||| ||||||||||||||| |
Db  910 GGTGTTGGATCCTCGGCAGAGGGGGCAGCCACCGCGCCGCAGGGGGCAGCCACCGG 969

Qy  678 CGGCGGGCCCCTCCTCCACGCTGCTGTGCTGCTCTTCCTTGTCTTTGGGCCCCATAGTCCCATGTCGT 737
           |  | ||||||||  ||| ||||
Db  970 GGCGGCCCACGCTGCTCAACTTCCTTGTCTTTGTGGGCCCTACAACGTGTCCCATGTCCT 1029

Qy  738 CAGCTTCCTGTAC 750
           || ||
Db 1030 GGGCTATATCTGC 1042
```

RESULT 5
US-09-248-715-1
; Sequence 1, Application US/09248715
Patent No. 6277960
GENERAL INFORMATION:
APPLICANT: BERGSMA, DERK
SATHE, GANESH M.
FUETTERER, WENDY
MAO, JOYCE
TITLE OF INVENTION: cDNA CLONE HNFDY20 THAT ENCODES
A NOVEL HUMAN 7-TRANSMEMBRANE RECEPTOR
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ratner & Prestia
STREET: P.O. Box 980
CITY: Valley Forge
STATE: PA
COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:

Qy 618 CCGGGCACTGGCCGCCGTCCGGCTGACGCACAGGCGGAAGCTGCGGCCGCCTGGGGTGGC 677
Db 621 GGTGTGGATCCTCGGCGAGAGGGGGCAGCCACCGGCGTAGGACCCTACAAGCGTGGC 680

Qy 678 CGGCGGGGCCCTCCTCACGCTGCTGCTCTCGCTGCTTGGGCCCTACAACGTGGC 737
Db 681 GGCGGCCACCGCTGCTCAACTTCCTTGTCTTGCTTTGGGCCCTACACGTGTCCCATGTCGT 740

Qy 738 CAGCTTCCTGTAC 750
Db 741 GGGCTATATCTGC 753

RESULT 7
US-09-170-496D-273
; Sequence 273, Application US/09170496D
; Patent No. 6555339
; GENERAL INFORMATION:
; APPLICANT: Behan, Dominic P.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: No. 6555339-Endogenous, Constitutively Activated Human G Protein-
; TITLE OF INVENTION: Receptors
; FILE REFERENCE: AREN-0040
; CURRENT APPLICATION NUMBER: US/09/170,496D
; CURRENT FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 273
; LENGTH: 1041
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-170-496D-273

Query Match   13.1%;   Score 118.6; DB 4;   Length 1041;
Best Local Similarity 51.8%;   Pred. No. 4.8e-16;
Matches 380;   Conservative 0; Mismatches 329; Indels 24; Gaps 4;

Qy   18 GCTCTCCTTCGGCCTCTATGTGGCCGCCTTCCCGCTTCAAGTCTCTGGC 77
Db   45 GTTCGTCTTCTCGGGTGTACCTTCTCACTTTCCTGGTGGGGCTCCCCCCACCTGCTGGC 104

Qy   78 CATCCGAGGGCGGACGGCCCACGCCCGGCTTCCGTCCTCACCCGTGGCTGGTCTAGCGCCCT 137
Db 105 CCTGGTGGTCTTCGTGGGCAAGCTGCAGGCGGTGGAGGC 164

Qy 138 GAACCTGGGCTGCTCCGACCTGCTGCTGCTGCCCTGAACAGTCTCTGCCCTGAAGGCGGTGAGGC 197
Db 165 CAACCTGACCGGCCTCGGACCTGCTCGCTGTTCCGCATGGGCATGGTGGAGGC 224

Qy 198 GCTAGCCTCCGGGGGCCTGGCCTGCCGGCCTTCTGCGGGTGTTCCGCGGTGGCCCA 257
Db 225 AGCCAATGGCATGCACTGGGCCCTTCATCCTCTGCCCACTCTTCTGGGATTCATCTT 284

Qy 258 CTTCTTCCCACTCTATGCGGGGGGCTTCCTGCGGCGCCCCTGAGTGCAGGCCGCTACCT 317
Db 285 CTTCACCACCATCATCTCACCGCCCTTCATCCTCTGGCAGCTGTGAGCATTGAACGCTTCCT 344

Qy 318 GGGAGCAGCCTTCCCCTTGGGCTACCAAGCTTCCGGAGGCGGTGCTATTCCTGGGGGGT 377
Db 345 GAGTGTGGCCCATCTGGGTCCGTGTGCCCGTCGGTCTTGGGGTCTTTTGGGTTGGAGGC 404

Qy 378 GTGCGCGGCCGCCATCTGGGCCCTCGTGTCACCTGGGTCTTTTGGGTTGGAGGC 437
Db 405 GAGTGTGGCCCATCTGGGTCCGTGTGCCCGTCGGTCTTGGGGTCTTTTGGGTTGGAGGC 464

Qy 438 TCCAGGAGGCGTGGCTGGCTGGACCACAGCAAACACCTCCCTGGGCATCAACACAACCGGTCAACGG 497
Db 465 ACATCTCCCACAGCCAGGGCACCAATGGGACCTG 506

Qy 498 CTCTCCGGTCGTCTGCCTGGAGGCCTGGACCCCGGCGCCCGCCTTCAGCCT 557
Db 507 CTACCTGGGAGTTCCGGAAGGACCAG--CTAGCCATCCTCCTGCCCGGCCTGGAG-AT 563

RESULT 6
US-09-170-496D-253

Db 970 GGCGGCCACCGCTGCTCAACTTCCTTGTCTTGCTTTGGGCCTTACACGTGTCCCATGTCGT 1029
Qy 738 CAGCTTCCTGTAC 750
Db 1030 GGGCTATATCTGC 1042

RESULT 6
US-09-170-496D-253
; Sequence 253, Application US/09170496D
; Patent No. 6555339
; GENERAL INFORMATION:
; APPLICANT: Behan, Dominic P.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: No. 6555339-Endogenous, Constitutively Activated Human G Protein-
; TITLE OF INVENTION: Receptors
; FILE REFERENCE: AREN-0040
; CURRENT APPLICATION NUMBER: US/09/170,496D
; CURRENT FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 253
; LENGTH: 1041
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-170-496D-253

Query Match   13.5%;   Score 121.8; DB 4;   Length 1041;
Best Local Similarity 52.1%;   Pred. No. 1e-16;
Matches 382;   Conservative 0; Mismatches 327; Indels 24; Gaps 4;

Qy   18 GCTCTCCTTCGGCCTCTATGTGGCCGCCTTCCCGCTTCAAGTCTCTGGC 77
Db   45 GTTCGTCTTCTCGGGTGTACCTTCTCACTTTCCTGGTGGGGCTCCCCCCACCTGCTGGC 104

Qy   78 CATCCGAGGGCGGACGGCCCACGCCCGGCTTCCGTCCTCACCCGTGGCTACGCGCCCT 137
Db 105 CCTGGTGGTCTTCGTGGGCAAGCTGCAGGCGGTGGAGGC 164

Qy 138 GAACCTGGGCTGCTCCGACCTGCTGCTGTTCCGCGGTGGAGGC 197
Db 165 CAACCTGACCGGCCTCGGACCTGCTCGCTGTTCCGCATGGGCATGGTGGAGGC 224

Qy 198 GCTAGCCTCCGGGGGCCTGGCCTGCCGGCCTCTGCGTGTCCGCGGTGGCCCA 257
Db 225 AGCCAATGGCATGCACTGGGCCCTTCATCCTCTGCCCACTCTTCTGGGATTCATCTT 284

Qy 258 CTTCTTCCCACTCTATGCGGGGGGCTTCCTGCGGCCGCCCTGAGTGCAGGCCGCTACCT 317
Db 285 CTTCACCACCATCATCTCACCGCCCTTCATCCTCTGGCAGCTGTGAGCATTGAACGCTTCCT 344

Qy 318 GGGAGCAGCCTTCCCCTTGGGCTACCAAGCTTCCGGAGGCGGTGCTATTCCTGGGGGGT 377
Db 345 GAGTGTGGCCCACCACTGTGGTACAAGACCCGGCCGAGGCGGGTCAGGCTCGGGTCTGGT 404

Qy 378 GTGCGCGGCCGCCATCTGGGCCCTCGTGTCACCTGGGTCTTTTGGGTTGGAGGC 437
Db 405 GAGTGTGGCCCACCACTGTGGTCTGTTGGCCTCTGCTCATCGTGCTCTCACTCATAGAAATT 464

Qy 438 TCCAGGAGGCGTGGCTGGCTGGACCACAGCAAACACCTCCCTGGGCATCAACACAACCGGTCAACGG 497
Db 465 CTCAGGGG------------ACATCTCCCACAGCCAGGGCACCAATGGGACCTG 506

Qy 498 CTCTCCGGTCGTCTGCCTGGAGGCCTGGACCCCGGCGCCCGCCTTCAGCCT 557
Db 507 CTACCTGGAGTTCCGGAAGGACCAG--CTAGCCATCCTCCTGCCCGGCCTGGAG-AT 563

Qy 558 CTCTCTCCTGCTCTTTTTCTGCCTCCATCACAGCCTTCTGCTACGTGGGCTGCCT 617
Db 564 GGCTGTGGTGGTCCTCTTTGTGTCCGCTGTCGCGCTGATCATCACCACCAGCTACGTCGTAC---AGCCCGCCT 620

Qy   558 CTCTCTCCTGCTCTTTTTCTGCCCTTGGCCATCACAGCCTTCTGCTACGTGGGCTGCCT 617
         ||| ||| ||||||| ||||||||||||| ||||||
Db   564 GGCTGTGGTCTCTTTGTCGTCCCGCTGATCATCACCAGCTACTGCTAC---AGCCGCCT 620

Qy   618 CCGGGCACTGGCCCCGCTCCGGGCTGACCGCACAGCGCGGAAGCTGCGGGCCTGGGTGGC 677
         ||||| || ||||| ||||||||| |||| ||||||||||||
Db   621 GGTGTGGATCCTGGCAGAGGGGCAGCCACCGCCGCGAGAGAGGGTGAAGGGGCTGTT 680

Qy   678 CGGGGGGCCCTCCACGCTGCTGTCTGCGTAGGACCCTACAACGCCTCCAACGTGCC 737
         || | ||||||| ||||||||||| || || ||||||||||||||||
Db   681 GGGCGCCCAGCTGCTCAACTTCCTTGTCTGCTTTGGGCCCTACACAACGTGTCCCATGTCGT 740

Qy   738 CAGCTTCCTGTAC 750
         ||||||||||
Db   741 GGGCTATATCTGC 753

RESULT 8
US-09-053-866-1
; Sequence 1, Application US/09053866
; Patent No. 6111075
; GENERAL INFORMATION:
; APPLICANT: Xu, Wenfeng
; APPLICANT: Presnell, Scott R.
; APPLICANT: Yee, David P.
; APPLICANT: Foster, Donald C.
; TITLE OF INVENTION: PROTEASE-ACTIVATED RECEPTOR
; TITLE OF INVENTION: PAR4 (ZCHEMR2)
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: ZymoGenetics, Inc.
;   STREET: 1201 Eastlake Avenue East
;   CITY: Seattle
;   STATE: WA
;   COUNTRY: USA
;   ZIP: 98102
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/053,866
;   FILING DATE:
;   CLASSIFICATION:
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
; ATTORNEY/AGENT INFORMATION:
;   NAME: Leith, Debra K
;   REGISTRATION NUMBER: 32,619
;   REFERENCE/DOCKET NUMBER: 98-10
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 206-442-6674
;   TELEFAX: 206-442-6678
;   TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 4895 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; FEATURE:
;   NAME/KEY: Coding Sequence
;   LOCATION: 176...1330
; OTHER INFORMATION:
US-09-053-866-1

Query Match   9.4%; Score 84.6; DB 3; Length 4895;
Best Local Similarity 47.2%; Pred. No. 5.4e-09;
Matches 329; Conservative   0; Mismatches 359; Indels 9; Gaps 2;

Qy   8 TGCCCCGCAGCTCTTCCTTGGGCCTTATGTGGCGGCCTTGCGTCGGGCTTCCCGCTCA 67
         |||| ||| ||||||||||| |||||||||||| |||| ||||||
Db   399 TGCCCACCAGGCTGGTGCCCGGCCCCTATGGGCTGGTGTCCTGGTGGTGGGCTGCCGGCCA 458

Qy   68 ACGTCCTGGCCATCCGAGGCGCGACGGCCCACGCGCCCGGCTCCGTCTCCGCTCACCCTGG 127
         | |||||||||||||||| |||||||| ||| |||||||||||||||||||
Db   459 ATGGGCTGGTCTGTGGTGCTGGCACCGCAGCAGGCACCTGCTGCTCCGGCTGCCACCATGTGC 518

Qy   128 TCTACGGCCTGAACCTGGCTGCTCCGGAGCTGCTGCTGACAGTCTCTCTGCCCCTGAAGG 187
             ||||||||||| ||| ||| |||||||||||||||||||||
Db   519 T------GATGAACCTCGGACTGCTGACCCTGGCCTCCAACGCCTCCAACGTGCC 572

Qy   188 CGGTGGAGGGCTAGCCTGCGGGCCTGGCCTCTGCGGCCTCGCCGTGCCCCGTCTTCG 247
         ||| ||||||||||||| ||| ||||||||||||||| |||| |||
Db   573 TCGCCTACCACCTGGCGTGGCCAGGCCGGTGGCCGTCGGGGAGGAGGCCGCCGCCTGCCTCGGCCA 632

Qy   248 CGGTGGCCCACTTCTTCCACTCTATGCCCGGGGGGGCTTCCTGGCCGCCGCCGCCTGAGTGCAG 307
         |||||| | |||| || ||||||||||||||| |||||||| ||| |||||||
Db   633 CGGCCGCACTCTATGGTCACATGTATGGCTCAGTGTCGTGCTGCTGGGCGCGCCGCCGCCTGG 692

Qy   308 GCCGCTACCTGGGAGCAGCAGCCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCGTGCTATT 367
         || ||| |||||| ||||||||||||||| |||| ||| |||||
Db   693 ATCGCTACCTGGCCCTGGTGCACCCGCTGCGGGCCCTGTCTCTGTCACGTGGCTGTCTTTG 752

Qy   368 CCTGGGGGGTGTGCGGCGCCATCTGGGCCTCGGCGCCCCTGTGTCACCTGGTCTGGTCTTTG 427
         |||| ||||||||||||| ||| | || ||||| ||
Db   753 CCCTTGGACTCTGCATGGCTGCTGG----CTCATCGCGGCGCCCTGGCACTGCCCTGA 809

Qy   428 GGTTGGAGGCTCCAGGAGGCTGGCTGGACCACAGCAACACACCTCCCTGGGCATCACACAC 487
         || ||||||||||||||||||||||| ||||||||||||||||||||
Db   810 CACTGCAGCGGCAGACACTTCCGGCTGGCGCGCTCCGATCGCGTGCTCTGCCATGACGCGC 869

Qy   488 CGGTCAACGGCTCTCCGGTCTGCCTGGAGCGCTGGAGACCACTGGGACCGGCCCTCGCCCGGCCC 547
         ||||||||||||||||||| ||||||||||||||||||||||||||||
Db   870 TGCCCCTGGACGCACAGGCGCTCCCCACTGGCAACCGGCTTCACCTGCTGCTGGCGCGTGTTGG 929

Qy   548 GCTTCAGCCTTCTCTCTCCTCTTTTTCTGCCCTTGGCCATCACAGCCTTCTCTGCTACG 607
         ||||||||| | || ||||||||||||||| |||||||||||||
Db   930 GCTGTTTCCTGCCCCTCGTGCCCATGCTGGCCATCTGTGTGCTACGGGCCACCGCACACGCTGG 989

Qy   608 TGGGCTGCCTCCCGGGCACTGGCCCCGCTCCGGGCTGACGCACAGGCGGAAGCTGCGGGCCG 667
         ||||||||||||||| ||||||||||||||||||||||||
Db   990 CGGCCAGCGGCCGGCGGCGCTACGGCGCCACGCGCGCTGAGGCTGACGGCGTGCTGGGCCTCCG 1049

Qy   668 CCTGGGTGGCCGGCGGGGGGCCCTCCTCACGCTGCT 704
         ||||||||||||||| |||||||||||
Db 1050 CCGTGGCCTTCTTCTTGTGCCCAGCAACCTGCTGCTGCT 1086

RESULT 9
US-09-479-130-1
; Sequence 1, Application US/09479130
; Patent No. 6436400
; GENERAL INFORMATION:
; APPLICANT: Xu, Wenfeng
; APPLICANT: Presnell, Scott R.
; APPLICANT: Yee, David P.
; APPLICANT: Foster, Donald C.
; TITLE OF INVENTION: PROTEASE-ACTIVATED RECEPTOR
; TITLE OF INVENTION: PAR4 (ZCHEMR2)
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: ZymoGenetics, Inc.
;   STREET: 1201 Eastlake Avenue East
;   CITY: Seattle
;   STATE: WA
;   COUNTRY: USA
;   ZIP: 98102
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/479,130
;   FILING DATE:
;   CLASSIFICATION:
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 206-442-6674
;   TELEFAX: 206-442-6678
;   TELEX:
; ATTORNEY/AGENT INFORMATION:
;   NAME: Leith, Debra K
;   REGISTRATION NUMBER: 32,619
;   REFERENCE/DOCKET NUMBER: 98-10
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 4895 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; FEATURE:
;   NAME/KEY: Coding Sequence
;   LOCATION: 176...1330
; OTHER INFORMATION:
US-09-479-130-1

Query Match   9.4%;   Score 84.6;   DB 4;   Length 4895;
Best Local Similarity 47.2%;   Pred. No. 5.4e-09;
Matches 329;   Conservative 0;   Mismatches 359;   Indels 9;   Gaps 2;

Qy   8 TGCCCCCGCAGCTCTCCTTGCCCTCTATGTGGCGCCTTTGCCGGCTTCCCGCTCA   67
Db   399 TGCCCACCAGGCTGGTGCCCGCCCTCTATGGCGCTGGTCCTGGTGGTGGGGGCTGCCGGCCA   458

Qy   68 ACGTCCTGGCCATCGAGGGCGACGGCCCACGCGCCGGGCTCCGTCTCACCCTAGCCTGG   127
Db   459 ATGGGCTGGCGCTGTGGGGTGCTGGCACCTCGCTGGCGCACCTCGCTGCCTCCACCATGCTGC   518

Qy   128 TCTACGCCCTGAACCTGGCTGCTCCGACAGTCTCTGCCCCTGAAGG   187
Db   519 T-----GATGAACCTCGGACTGCTGACCTCCTGGGCCTGCCCCCGGGA   572

Qy   188 CGGTGGAGGGCTAGCCTCCGGGGGCGCTTCCTGGCCGCCGTGAGTGCAG   247
Db   573 TCGCCTACCACCTGCGTGGCCAGCCGCCTTGGGGAGGCCGCCTGGCCTCAGCCTGG   632

Qy   248 CGGTGGCCCACTTCTTCCCACTCTATGCGCGGCGGGGCTTCCTGGCCGCCGTGAGTGCAG   307
Db   633 CGGCCGCACTCTATGGTCACATGTATGGCTCAGTGCTGCTGCTGGCCGCCGTCAGCCTGG   692

Qy   308 GCCGCTACCTGGGAGCAGCAGCCTTCCCCTTGGGCTACCAAGCCTTCGGAGGCCGTGCTATT   367
Db   693 ATCGCTACCTGGCCCGTTGCACCCTGGGGGGCCCGCTCCTGTGTCACCGGGTCTGGTCTTTG   752

Qy   368 CCTGGGGGGTGTGCGGCGCCATCGGGCCAGGAGGCTGGCTCGTGTCCTGGTCTCGTCTTTG   427
Db   753 CCCTTGGACTCTGCATGGCTGCTTGG---CTCATGCGGCGGCCTGCTGGGGTCGGCGGTCTTTG   809

Qy   428 GGTTGGAGGCTCCAGGAGGCTGGCTGGACCACAGCATCAACACAC   487
Db   810 CACTGCAGCGGCAGACCTTCGGCTCTGCCATGCTCTGCCATGACGCGC   869

Qy   488 CGGTCAACGGCTCTCCGGTTCGCCTGGAGGCCTGGGACCCGGGCCC   547
Db   870 TGCCCCTGGACGCACAGGCCTCCCCACTGGCAACCGGCCTTCACCTGCCCTGGGCGTGTTGG   929

Qy   548 GCTTCAGCCTCTCTCTGCTCTTTTTCTGCCCCTTGCCATCACAGCCTTCTGCTACG   607
Db   930 GCTGTTTCCTGCCCGCTGCCATGCTGCGTGTCGTGCTACGGGGCCACCGCATGCTGG   989

Qy   608 TGGGCTGCCTCCGGGCACTGCTGCCCCGCTCCGGGCGCTGGACTGGAAGCTGGCGGGGCG   667

（right column）

Db   990 CGGCCAGCGGCCGGGGCTACGGCCACGCGCTGAGGCTGACCGCAGTGGTGCTGGCCTCCG 1049
Qy   668 CCTGGGTGGCCGCGGGCCCTCCTCACGCTGCTGCT   704
Db 1050 CCGTGCGGCCTTCTTCGTGCCCAGCAACCTGCTGCT 1086

RESULT 10
US-09-472-130A-1
; Sequence 1, Application US/09472130A
; Patent No. 6473765
; GENERAL INFORMATION:
; APPLICANT: Xu, Wenfeng
; APPLICANT: Presnell, Scott R.
; APPLICANT: Yee, David P.
; APPLICANT: Foster, Donald C.
; TITLE OF INVENTION: PROTEASE-ACTIVATED RECEPTOR PAR4
; TITLE OF INVENTION: (ZCHEMR2)
; FILE REFERENCE: 98-10D2
; CURRENT APPLICATION NUMBER: US/09/472,130A
; CURRENT FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/053,866
; PRIOR FILING DATE: 1998-04-01
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 4895
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (176)...(1330)
US-09-472-130A-1

Query Match   9.4%;   Score 84.6;   DB 4;   Length 4895;
Best Local Similarity 47.2%;   Pred. No. 5.4e-09;
Matches 329;   Conservative 0;   Mismatches 359;   Indels 9;   Gaps 2;

Qy   8 TGCCCCCGCAGCTCTCCTTGCCCTCTATGTGGCGCCTTTGCCGGCTTCCCGCTCA   67
Db   399 TGCCCACCAGGCTGGTGCCCGCCCTCTATGGCGCTGGTCCTGGTGGTGGGGGCTGCCGGCCA   458

Qy   68 ACGTCCTGGCCATCGAGGGCGACGGCCCACGCGCCGGGCTCCGTCTCACCCTAGCCTGG   127
Db   459 ATGGGCTGGCGCTGTGGGGTGCTGGCACCTCGCTGGCGCACCTCGCTGCCTCCACCATGCTGC   518

Qy   128 TCTACGCCCTGAACCTGGCTGCTCCGACAGTCTCTGCTGCCCCTGAAGG   187
Db   519 T-----GATGAACCTCGGACTGCTGACCTCCTGGGCCTGCCCCCGGGA   572

Qy   188 CGGTGGAGGGCTAGCCTCCGGGGGCTGCCTGGCCGCCGTGCGTGCCCGTCTTCG   247
Db   573 TCGCCTACCACCTGCGTGGCCAGCCGCCTTGGGGAGGGCCGCCTGGCCTCAGCCTGG   632

Qy   248 CGGTGGCCCAGTTCTTCCCACTCTATGCGGCGGGGCTTCCTGGCCGCCGAGTCGCAG   307
Db   633 CGGGCCGCACTCTATGGTCACATGTATGGCTCAGTGCTGCTGCTGGCCGCCGTCAGCCTGG   692

Qy   308 GCCGTACCTGGGGGACGCAGCAGCCTTCCCCTTGGGCTACCAAGCCTTCGGAGGCCGTGTATT   367
Db   693 ATCGCTACCTGGCCCGTTGCACCCTGGGCCGCATCGGTGCTCTGCCATGCGCGCGGCCTGG   752

Qy   368 CCTGGGGGGTGTGCGGCGCCATCGGGCCATCCTGGGACCCGGGCCCTCGGTGGTCTTTG   427
Db   753 CCCTTGGACTCTGCGATGGCTGCTTGG---CTCATGCGGCGGCCTGCTGGCACTGCCCCCTGA   809

Qy   428 GGTTGGAGGCTCCAGGAGGCTGGCTGGACCACAGCAACACCTCCCTGGGCATCAACACAC   487
Db   810 CACTGCAGCGGCAGACCTTCGGCTCTGCCATGCTCTGCCATGACGCGC   869

Qy   488 CGGTCAACGGCTCTCCGGTCGCTGCTGCCACTGGTGCTGCGGGGCCTTCGCCGGCCCC   547
Db   870 TGCCCCTGGACGCACAGGCCTCCCCACTGGCAACCGGCCTTCACCTGCCCTGCGCGCGGTGTTGG   929

; APPLICANT: Behan, Dominic P.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: No. 6555339-Endogenous, Constitutively Activated Human G Protein-C
; TITLE OF INVENTION: Receptors
; FILE REFERENCE: AREN-0040
; CURRENT APPLICATION NUMBER: US/09/170,496D
; CURRENT FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 225
; LENGTH: 1098
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-170-496D-225

Query Match   8.9%;   Score 80; DB 4; Length 1098;
Best Local Similarity 51.8%;   Pred. No. 4.2e-08;
Matches 207;   Conservative   0; Mismatches 190;   Indels   3;   Gaps   1;

Qy   6 CCTGCCCCCGCCAGCTCTCCTTCGGCCTCTCATGTGGCCGCCTTGCGCTGGCTTCCCGCT 65
            ||| | |||| ||||||||| |||||||||||||||||||||||||||||
Db   54 CATCCACCAGACGCTGGCCCCGGTGGTCTATGTTACCGTGGCTGGTGGGCTTCCCGGC 113

Qy   66 CAACGTCCTGGCCATCCGAGGCGCGACGGCCCACGCCCCGGCTCCGCCCTCACCCCTAGCCT 125
            |||||| ||||||||||||||||||||||||||||||||||||||||||
Db   114 CAACTGCCTGTCCCTCTACTTCGGCTACCTGGCTGCTGCGATCAAGGCCCGGAACGAGCTGGGC-- 171

Qy   126 GGTCTACGCCCTGAACCTGGCTGCTCCGACCTGCTGTGCTGACAGTCTCTCGCCCCTGAA 185
            ||| |||||| ||||||||||||||||||||||||||||||||||
Db   172 -GTGTACCTGTGCAACCTGACGCGTGGCCGACCTCTTCTACATCTGCTCGCTGCGCCCTTCTG 230

Qy   186 GGCGGTGGAGGCGCTAGCCTCCGGGGGCCTCCGGCGGGGGCTTCCTGGCCCGCCCGTCTT 245
            |||||||||||||| ||||||||||||||||||||||||||||||||
Db   231 GCTGCAGTACGTGCTGCAGCACGACAACTTGCTCTCACGGCGACCTGGTCCTGCCAGGTGTG 290

Qy   246 CGCGGTGGCCCACTTCTTCCTGGCCATCTATGTGGGCGGGGGCTTCCTGGAGGCCGTGCTA 305
            ||| ||| |||||||||||||||||||||||||||||||||||||||
Db   291 CGGCATCCTCCTGACGAGAACATCTACATCAGCCTGTGGCTTCCTGCTGCATCTCCGT 350

Qy   306 AGGCCGCTACCTGGGAGCAGCCTTCCCCCTTGGGCCGCCATCGGCCTCCTG 405
            |||||||||||||||||||||||||||||
Db   351 GGACCGCTACCTGGCTGTGCCCATCCCTTCCGGCTTCCACCAGTTCGGACCCTGAAGGC 410

Qy   366 TTCCTGGGGGTGTGGCGGCCCATCTGGGCCCTCCTG 405
            |||| |||||| ||||||||||||||||||
Db   411 GGCCGTCGGCGTCAGCGTGGTCATCTGGGCCAAGGAGCTG 450

RESULT 13
US-08-724-974A-1
; Sequence 1, Application US/08724974A
; Patent No. 5912335
; GENERAL INFORMATION:
; APPLICANT: Derk J. Bergsma, Catherine E. Ellis
; TITLE OF INVENTION: A No. 5912335el G-Protein Coupled Receptor
; TITLE OF INVENTION: HUVCT36
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road, P.O. Box 1539
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
; COMPUTER: IBM 486
; OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,974A
; FILING DATE: October 3, 1996

Qy   548 GCTTCAGCCTCTCCTGCTCTTTTTTCTGCCCTTGGCCATCACAGCCTTCTGCTACG 607
            |||||||||||||||||||||||||||||||||||||||||||||||
Db   930 GCTGTTTCCTGCCCCTGCTGGCCATGCTGCTGTGCTACGGGCCACCCTGCACACGCTGG 989

Qy   608 TGGGCTGCTCCGGCACTGGCCTGCTTCGGCGCCCACTGGCCTGACGCGAAGCTGCGGGGCCG 667
            |||||||||| |||||
Db   990 CGGCCAGCGGCCGGCGCTACGGCGCACGCGGCTGACGCCGAGTGGTGCTGGCCTCCG 1049

Qy   668 CCTGGGTGCCGGCGGCGGGCCTCCTCACGCTGCTGCT 704
            |||||||||| |||||||||||
Db   1050 CCGTGGCCTTCTTCGTGCCGCCAGCAACCTGCTGCTGCT 1086

RESULT 11
US-09-170-496D-117
; Sequence 117, Application US/09170496D
; Patent No. 655339
; GENERAL INFORMATION:
; APPLICANT: Behan, Dominic P.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: No. 6555339-Endogenous, Constitutively Activated Human G Protein-
; TITLE OF INVENTION: Receptors
; FILE REFERENCE: AREN-0040
; CURRENT APPLICATION NUMBER: US/09/170,496D
; CURRENT FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 117
; LENGTH: 1098
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-170-496D-117

Query Match   8.9%;   Score 80; DB 4; Length 1098;
Best Local Similarity 51.8%;   Pred. No. 4.2e-08;
Matches 207;   Conservative   0; Mismatches 190;   Indels   3;   Gaps   1;

Qy   6 CCTGCCCCCGCCAGCTCTCCTTCGGCCTCTCATGTGGCCGCCTTTGCGCTGGCTTCCCGCT 65
            ||| | |||| ||||||||| ||||||||||||||||||||||||
Db   54 CATCCACCAGACGCTGGCCCCGGTGGTCTATGTTACCGTGGCTGGTGGGCTTCCCGGC 113

Qy   66 CAACGTCCTGGCCATCCGAGGCGCGACGGCCCACGCCCCGGCTCCGCCCTCACCCCTAGCCT 125
            ||||||||||||||||||||||||||||||||||||
Db   114 CAACTGCCTGTCCCTCTACTTCGGCTACCTGGCTGCTGCGATCAAGGCCCGGAACGAGCTGGGC-- 171

Qy   126 GGTCTACGCCCTGAACCTGGCTGCTCCGACCTGCTGTGCTGACAGTCTCTCGCCCCTGAA 185
            ||| |||||| ||||||||||||||||||||||||||||||
Db   172 -GTGTACCTGTGCAACCTGACGCGTGGCCGACCTCTTCTACATCTGCTCGCTGCGCCCTTCTG 230

Qy   186 GGGCGGTGGAGGCGCTAGCCTCCGGGGGCCTCCGGCGGGGGCTTCCTGGCCCGCCCGTGTG 245
            |||||||||||||| |||||||||||||||||||||||||||
Db   231 GCTGCAGTACGTGCTGCAGCAGCACGACAACTTGCTCTCACGGCGACCTGGTCCTGCCAGGTGTG 290

Qy   246 CGCGGTGGCCCACTTCTTCGGCCTCTCTATGTGGCCGCCGGGGGCTTCCTGGGCCTTTGCGGGC 305
            ||| || | ||| |||||||||||||||||||||
Db   291 CGGCATCCTCCTGACGAGAACATCTACATCAGCCTGTGGCTGCTCCGGCTTCCTGCTGCATCTCCGT 350

Qy   306 AGGCCGCTACCTGGGAGCAGCCTTCCCCCTTGGGCCGCCTACCAAGGCCGGTGCTA 365
            ||||||||||||||||||||||||||
Db   351 GGACCGCTACCTGGCTGTGCCCATCCCTTCCGGCTTCCACCAGTTCGGACCCTGAAGGC 410

Qy   366 TTCCTGGGGGTGTGCGCGGCCCATCTGGGCCCTCCTG 405
            |||| ||||||| ||||||||||||||
Db   411 GGCCGTCGGCGTCAGCGTGGTCATCTGGGCCAAGGAGCTG 450

RESULT 12
US-09-170-496D-225
; Sequence 225, Application US/09170496D
; Patent No. 655339
; GENERAL INFORMATION:

```
;    CLASSIFICATION: 424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME: William T. Han
;      REGISTRATION NUMBER: 34,344
;      REFERENCE/DOCKET NUMBER: ATG50022
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: 610 270 5024
;      TELEFAX: 610 270 5090
;  INFORMATION FOR SEQ ID NO: 1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 1597
;      TYPE: Nucleic Acid
;      STRANDEDNESS: Single
;      TOPOLOGY: Linear
;      ANTI-SENSE: No
US-08-724-974A-1

Query Match            8.9%;   Score 80;  DB 2;  Length 1597;
Best Local Similarity 51.8%;   Pred. No. 4.3e-08;
Matches  207;  Conservative   0;  Mismatches 190;  Indels   3;  Gaps   1;

Qy      6 CCTGCCCCCGCAGCTCTCCTTCGGCCTCTATGTGGCCGCCTTTGCGCTGGGCTTCCCGCT 65
            |  | || || ||||    |||||    ||||||| | |||| | |||||||||||| |
Db    395 CATCCACCAGACGCTGGCCCCGGTGGTCTATGTTACCGTGCTGGTGGTGGTGGGCTTCCCGGC 454

Qy     66 CAACGTCCTGGCCATCCAGAGGCGCGACGGCCCACGCCCGGCTCCGTCACCCCTAGCCT 125
            |||| ||||| |   ||       |||  || | || ||  |||  | |||  |||| |
Db    455 CAACTGCCTGTCCCCTCACTTCGGCTACCTGCAGATCAAGGCCCGGAACGAGCTGGGC-- 512

Qy    126 GGTCTACGCCCTGAACCTGGCTGCTCCGACCTGTCTCTGCCCCTGAA 185
            | | |||   | ||| ||| |   |||||||| || | ||| | |
Db    513 -GTGTACCTGTGCAACCTGACGGTGCCGACCTCTTCTACATCTGCTCGCTGCCCTTCTG 571

Qy    186 GGCGGTGGAGGCGCTAGCCTCCGGGGCGTGGCCTTGCGCTCGCTGTGCCCCGTCTT 245
            |||  |||| |||  || | |||   || || ||    |   || |||  |||
Db    572 GCTGCAGTACGTGCGCAGCAGCACGACAAACTGGTCTCACGCGGACCGTGTCCCAGGTGTG 631

Qy    246 CGCGGTGGCCCACTTCTTCCCACTCTATGCCGGCGGGGGCTTCCTGCCGCCCTGAGTGC 305
            || |  | |||   |||  ||   || |  |  |||||||||||| ||  ||| |
Db    632 CGGCATCCTCCTGTACGAGACAACATCTACACTCAGCGTGGGCTTCCTCGTCGTGCCATCTCCGT 691

Qy    306 AGGCCGCTACCTGGAGGCAGCCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCGTGCTA 365
            ||| ||||||||||  | |  ||| ||| | | | |  |||||||||| | ||| |||
Db    692 GGACCGCTACCTGGCGTGTGCCCATCCCTTCGCTTCCACCAGTTCCGGACCCTGAAGGC 751

Qy    366 TTCCTGGGGGGTGTGCGCGGCCATCTGGGCCCCTCCTG 405
            | |||||||| | || ||| ||||||||| |  ||
Db    752 GGCCGTCGGCGTCAGCGTGGTCATCTGGGCGGCCAAGGAGCTG 791


RESULT 14
US-09-364-425B-26
; Sequence 26, Application US/09364425B
; Patent No. 6653086
; GENERAL INFORMATION:
;   APPLICANT: Behan, Dominic P.
;   APPLICANT: Chalmers, Derek T.
;   APPLICANT: Liaw, Chen W.
;   APPLICANT: Lin, I-Lin
;   APPLICANT: Lowitz, Kevin P.
;   APPLICANT: Chen, Ruoping
;   TITLE OF INVENTION: Endogenous,Constitutively Activated G Protein-Coupled Orphan Rece
;   FILE REFERENCE: Aren0047
;   CURRENT APPLICATION NUMBER: US/09/364,425B
;   CURRENT FILING DATE: 2001-12-18
;   PRIOR APPLICATION NUMBER: 60/094,879
;   PRIOR FILING DATE: 1998-07-31
;   PRIOR APPLICATION NUMBER: 60/106,300
;   PRIOR FILING DATE: 1998-10-30
;   PRIOR APPLICATION NUMBER: 60/110,906
```

```
;   PRIOR FILING DATE: 1998-12-04
;   PRIOR APPLICATION NUMBER: 60/121,851
;   PRIOR FILING DATE: 1999-02-26
;   NUMBER OF SEQ ID NOS: 60
;   SOFTWARE: Patentin version 3.1
SEQ ID NO 26
;   LENGTH: 1697
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-364-425B-26

Query Match            8.9%;   Score 80;  DB 4;  Length 1697;
Best Local Similarity 51.8%;   Pred. No. 4.3e-08;
Matches  207;  Conservative   0;  Mismatches 190;  Indels   3;  Gaps   1;

Qy      6 CCTGCCCCCGCAGCTCTCCTTCGGCCTCTATGTGGCCGCCTTTGCGCTGGGCTTCCCGCT 65
            |  | || || ||||    |||||    ||||||| | |||| | |||||||||||| |
Db    377 CATCCACCAGACGCTGGCCCCGGTGGTCTATGTTACCGTGCTGGTGGTGGTGGGCTTCCCGGC 436

Qy     66 CAACGTCCTGGCCATCCAGAGGCGCGACGGCCCACGCCCGGCTCCGTCACCCCTAGCCT 125
            |||| ||||| |   ||       |||  || | || ||  |||  | |||  |||| |
Db    437 CAACTGCCTGTCCCCTCACTTCGGCTACCTGCAGATCAAGGCCCGGAACGAGCTGGGC-- 494

Qy    126 GGTCTACGCCCTGAACCTGGCTGCTCCGACCTGTCTCTGCCCCTGAA 185
            | | |||   | ||| ||| |   |||||||| || | ||| | |
Db    495 -GTGTACCTGTGCAACCTGACGGTGCCGACCTCTTCTACATCTGCTCGCTGCCCTTCTG 553

Qy    186 GGCGGTGGAGGCGCTAGCCTCCGGGGCGTGGCCTTGCGCTCGCTGTGCCCCGTCTT 245
            |||  |||| |||  || | |||   || || ||    |   || |||  |||
Db    554 GCTGCAGTACGTGCGCAGCAGCACGACAACTGGTCTCACGCGGACCGTGTCCCAGGTGTG 613

Qy    246 CGCGGTGGCCCACTTCTTCCCACTCTATGCCGGCGGGGGCTTCCTGCCGCCCTGAGTGC 305
            || |  | |||   |||  ||   || |  |  |||||||||||| ||  ||| |
Db    614 CGGCATCCTCCTGTACGAGACAACATCTACACTCAGCGTGGGCTTCCTCGTCGTGCCATCTCCGT 673

Qy    306 AGGCCGCTACCTGGAGGCAGCCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCGTGCTA 365
            ||| ||||||||||  | |  ||| ||| | | | |  |||||||||| | ||| |||
Db    674 GGACCGCTACCTGGCGTGTGCCCATCCCTTCGCTTCCACCAGTTCCGGACCCTGAAGGC 733

Qy    366 TTCCTGGGGGGTGTGCGCGGCCATCTGGGCCCCTCCTG 405
            | |||||||| | || ||| ||||||||| |  ||
Db    734 GGCCGTCGGCGTCAGCGTGGTCATCTGGGCGGCCAAGGAGCTG 773


RESULT 15
US-09-762-661A-1
; Sequence 1, Application US/09762661A
; Patent No. 6645726
; GENERAL INFORMATION:
;   APPLICANT: Howard, Andrew D.
;   APPLICANT: Palyha, Oksana C.
;   APPLICANT: Smith, Roy G.
;   APPLICANT: Tan, Carina P.
;   TITLE OF INVENTION: CANINE GROWTH HORMONE SECRETAGOGUE
;   TITLE OF INVENTION: RECEPTOR
;   FILE REFERENCE: 20207P
;   CURRENT APPLICATION NUMBER: US/09/762,661A
;   CURRENT FILING DATE: 2001-05-30
;   PRIOR APPLICATION NUMBER: PCT/US99/17915
;   PRIOR FILING DATE: 1999-08-06
;   PRIOR APPLICATION NUMBER: 60/095,960
;   PRIOR FILING DATE: 1998-08-10
;   NUMBER OF SEQ ID NOS: 9
;   SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 1
;   LENGTH: 1050
;   TYPE: DNA
;   ORGANISM: Canis familiaris
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1050)
;   OTHER INFORMATION: n = A,T,C or G
US-09-762-661A-1
```

```
Query Match          8.7%;  Score 78.2;  DB 4;  Length 1050;
Best Local Similarity  47.9%;  Pred. No. 9.7e-08;
Matches  292;  Conservative   0;  Mismatches  309;  Indels    9;  Gaps    2;

Qy     28 GGCCTCTATGTGGCCGCCTTTGCGCTGGGCTTCCCGCTCAACGTCCTGGCCATCCGAGGGC  87
          |||| |||| || |||||  | ||| ||| |||| ||||| |||| |  | ||   |||
Db     91 GCCACCTGCGTGGCGCGCTGTTCGCGTGGGCGTGCGGCAACCTGCTGACGGTGCTGGTG 150

Qy     88 GCGACGGCCCACGCCCGGCTCCGTCTCTCACCCCTAGCCTGGTCTACGCCCTGAACCTGGGC 147
          ||||  || ||  ||  |||||||  | |||| |||||  ||  |||  ||| |||||
Db    151 GTGCGGCGGCGTTCCGCGAGCTGCGGCACCACCAACCTG---TACCTGTGCAGCCTGGCC 207

Qy    148 TGCTCCGACCTGTGCTGACAGTCTCTCTGCCCCTGAAGGGCGGTGGAGGCGCTAGCCTCC 207
          ||| |||||  |  || ||| || |  |||| |||  | ||  ||  ||| |  ||
Db    208 TGCTCCGACCTGCTCATCTTCCTGTGCATGCCGCTCGACCTGGTGCGCCTGTGGCAGTAC 267

Qy    208 GGGGCCCTGCCCTCTGCCGCCGCCGTGCCCGTGCCCGTCTTCGCGGTGGCCCACTTCTTCCCA 267
          ||| |||||| ||||| | ||    || ||  | |  || |||| ||  ||  |
Db    268 CGGCCCCTGGACCTTCGGCGACCTGCTCTGCAAACTCTTCCAGTTCGTGAGCGAGGGGCTGC 327

Qy    268 CTCTATGCCGGCGGGGGCTTCCTGGCCGCCCTGAGTGCAGGCCGCTACCTGGGAGCAGCC 327
          ||| |  ||| | |||||| | |||||  || || ||  |||  | |||| ||| || ||
Db    328 ACCTACGCCACGTGCTCACCATCACGGCGTCGAGCGTCGAGCGCTACTTCGCCATCTGC 387

Qy    328 TTCCCCTTGGGCTACCAAGCCTTCCGGAGGGCCGTGCTATTCCTGGGGGGTGTTGCGCGCC 387
          |||||  | ||  ||  ||  ||||  ||| || || || ||   ||| |  || | ||
Db    388 TTCCCGCTGCGCGCAAGGTGCTGGTGACCAAGGGCCGCGGTGAAGCTGCCCTGCTGGCC 447

Qy    388 ATCTGGGGCCCTGTCGTGTCACCTGGGTCTGTCTGTCTTTTGGGTTGGAGGCTCCAGGAGGC 447
          ||||||||| ||| | ||||  ||| |||| | |||| | || ||  ||||||  ||||
Db    448 ATCTGGGGCGTGGCCTTCTGCAGCGCGCCGGGACCCCATCTTCGTGTCGGTG------GGCGTG 501

Qy    448 TGGCTGGACCAGCAGCAACACCTCCCTGGGCATCAACACACACCGGTCAACGGCTCTCCGGTC 507
          ||||||||| ||| |  |||| ||  ||  || |||| | || ||  |||||| | ||||
Db    502 GAGCACGAGAACGGCACCACCCGGACCCCCGGGACACCCGGTCGCGCGCGCCACCGAGTTCGC 561

Qy    508 TGCCTGGAGGGCTTGGGACCCGGCGCCGGGCTCTGCGCGGCCGCCGCTTCAGCCTCTCTCCTG 567
          | ||| ||||  || ||    | |  || || ||  |||   || ||  |||| | ||||||
Db    562 GTGCGCTCGGGGCTGCTCACGGCCATGGTGTCCAGCGTGTCCAGCGTCTTCTTCTTTCCTGCCC 621

Qy    568 CTCTTTTTCTGCCCTTGGCCATCACAGCCTTCTGCTACGTGGGCTGCCTCCGGGCACTG 627
          |||  ||||||||||   ||||||||||||  | |||| ||  || ||||||  ||  ||
Db    622 GTCTTCTGCCTCACGGTGCTCTACGGCGCGCGAGAAGCTGTGGCGGGCGGGGGGC 681

Qy    628 GCCCGCTCCG 637
          |  | ||
Db    682 GGCGACACNG 691
```

Search completed: April 30, 2004, 15:43:30
Job time : 93 secs

Copyright   GenCore version 5.1.6
            Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2004, 15:43:35 ; Search time 465 Seconds
                                       (without alignments)
                                       8784.275 Million cell updates/sec

Title: US-10-202-687-1
Perfect score: 903
Sequence: 1 atggacctgccccgcagct..........ggggcaagtcccagagtaa 903

Scoring table: IDENTITY_NUC
            Gapop 10.0 , Gapext 1.0

Searched: 2936184 seqs, 226173022 residues

Total number of hits satisfying chosen parameters: 5872368

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Published Applications NA:*
    1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
    2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
    3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
    4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
    5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
    6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
    7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
    8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
    9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
   10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
   11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
   12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
   13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
   14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
   15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
   16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
   17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
   18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
   19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query % Match	Query Length	DB	ID	Description
1	903	100.0	903	15	US-10-251-385-249	Sequence 249, App
2	903	100.0	903	15	US-10-225-567A-474	Sequence 474, App
3	903	100.0	903	15	US-10-029-386-22904	Sequence 22904, A
4	903	100.0	903	16	US-10-202-687-1	Sequence 1, Appli
5	898.2	99.5	903	15	US-10-251-385-271	Sequence 271, App
6	154	17.1	550	15	US-10-029-386-9201	Sequence 9201, Ap
7	121.8	13.5	960	15	US-10-029-386-24088	Sequence 24088, A
8	121.8	13.5	1041	15	US-10-251-385-253	Sequence 253, App
9	121.8	13.5	1041	15	US-10-225-567A-604	Sequence 604, App
10	121.8	13.5	1041	15	US-10-203-539-1	Sequence 1, Appli
11	121.8	13.5	1041	15	US-10-203-539-3	Sequence 3, Appli
12	121.8	13.5	1206	15	US-10-029-386-25146	Sequence 25146, A
13	118.6	13.1	1041	15	US-10-251-385-273	Sequence 273, App
14	84.6	9.4	4895	15	US-10-225-567A-515	Sequence 515, App

15	84.6	9.4	4895	15	US-10-187-049-1	Sequence 1, Appli
16	80	8.9	1098	15	US-09-850-948-5	Sequence 5, Appli
17	80	8.9	1098	15	US-10-251-385-117	Sequence 117, App
18	80	8.9	1098	15	US-10-251-385-225	Sequence 225, App
19	80	8.9	1098	15	US-10-273-575-5	Sequence 5, Appli
20	80	8.9	1098	15	US-10-225-567A-282	Sequence 282, App
21	80	8.9	1434	15	US-10-029-386-25332	Sequence 25332, A
22	80	8.9	1697	15	US-10-109-533A-1	Sequence 1, Appli
23	80	8.7	1697	15	US-10-288-222A-11	Sequence 11, Appl
24	78.2	8.7	1329	9	US-09-853-161-25	Sequence 25, Appl
25	78.2	8.7	1329	9	US-09-852-659A-25	Sequence 25, Appl
26	78.2	8.7	1329	9	US-09-852-797-25	Sequence 25, Appl
27	78.2	8.7	1329	13	US-10-058-993-25	Sequence 25, Appl
28	77.2	8.5	993	15	US-10-251-385-257	Sequence 257, App
29	77.2	8.5	993	15	US-10-225-567A-466	Sequence 466, App
30	77.2	8.5	993	15	US-10-337-992-1	Sequence 1, Appli
31	77.2	8.5	994	15	US-10-029-386-25403	Sequence 25403, A
32	77.2	8.5	1616	15	US-10-348-190-3	Sequence 3, Appli
33	75.6	8.4	993	15	US-10-251-385-275	Sequence 275, App
34	73.2	8.1	2426	15	US-10-348-190-1	Sequence 1, Appli
35	70.8	7.8	960	15	US-10-203-539-5	Sequence 5, Appli
36	70.4	7.8	1955	15	US-10-400-991-2	Sequence 2, Appli
37	70.4	7.8	1955	15	US-10-190-469-2	Sequence 2, Appli
38	70	7.8	1116	15	US-10-225-567A-361	Sequence 361, App
39	70	7.8	2427	9	US-09-254-783A-2	Sequence 2, Appli
40	70	7.8	2427	14	US-10-152-058-2	Sequence 2, Appli
41	68.4	7.6	1080	13	US-09-875-076-27	Sequence 27, Appl
42	68.4	7.6	1080	13	US-09-876-252-29	Sequence 29, Appl
43	68.4	7.6	1080	15	US-10-272-983-27	Sequence 27, Appl
44	68.4	7.6	1080	15	US-10-393-807-27	Sequence 27, Appl
45	68.4	7.6	1080	16	US-10-417-820A-29	Sequence 29, Appl

ALIGNMENTS

RESULT 1
US-10-251-385-249
; Sequence 249, Application US/10251385
; Publication No. US20030105292A1
; GENERAL INFORMATION:
; APPLICANT: Behan, Dominic P.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: No. US20030105292A1-Endogenous, Constitutively Activated Human G
; TITLE OF INVENTION: Protein-Coupled
; TITLE OF INVENTION: Receptors
; FILE REFERENCE: AREN-0040
; CURRENT APPLICATION NUMBER: US/10/251,385
; CURRENT FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: US/09/170,496
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 249
; LENGTH: 903
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-251-385-249

Query Match   100.0%; Score 903; DB 15; Length 903;
Best Local Similarity   100.0%; Pred. No. 3.8e-210;
Matches 903; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 ATGGACCTGCCCCGCAGCTCTCCTTCGGCCTCTATGGTGGCCGCCGGGCTTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGGACCTGCCCCGCAGCTCTCCTTCGGCCTCTATGGTGGCCGCCGGGCTTC 60

Qy   61 CCGCTCAACGTCCTGGCCATCCAGCCGGACGCGGCCCACGCGGTCCGGTCTCACCCCT 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CCGCTCAACGTCCTGGCCATCCAGCCGGACGCGGCCCACGCGGTCCGGTCTCACCCCT 120

Qy   121 AGCCTGCTCTACGCCCTGAACCTGGCGTGCTCCGACCTGCTCTGACAGTCTTCTCGCCC 180

```
; SEQ ID NO 474
; LENGTH: 903
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-225-567A-474

Query Match          100.0%;    Score 903;  DB 15;    Length 903;
Best Local Similarity 100.0%;    Pred. No. 3.8e-210;
Matches 903;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy      1 ATGAACCTGCCCCGCAGTCTCTCCTTGGCCCTCTATGTGCCGCCTGGGCTTC 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGGAACCTGCCCCGCAGTCTCTCCTTGGCCCTCTATGTGCCGCCTGGGCTTC 60

Qy     61 CCGCTCAACGTCTGGCCATCCGAGGGCGACGGCCCAGCCCGCTCCGTCCACCCCT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCGCTCAACGTCTGGCCATCCGAGGGCGACGGCCCAGCCCGCTCCGTCCACCCCT 120

Qy    121 AGCCTGGTCACGCCCTGAACCTGGCTGCTCCGACCTGCTGCTGCTGACAGTCTCTGCCC 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AGCCTGGTCACGCCCTGAACCTGGCTGCTCCGACCTGCTGCTGCTGACAGTCTCTGCCC 180

Qy    181 CTGAAGGCCGTGGAGGCGCTAGCCTGTGCCCCTGGCTAGCCTTCCGGAGGCCG 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CTGAAGGCCGTGGAGGCGCTAGCCTGTGCCCCTGGCTAGCCTTCCGGAGGCCG 240

Qy    241 GTCTTCGCGGTGGCCCACTCTTCCCACTCTATGCCGGGGGCTTCTGGCCGCCCCTG 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GTCTTCGCGGTGGCCCACTCTTCCCACTCTATGCCGGGGGCTTCTGGCCGCCCCTG 300

Qy    301 AGTGCAGGCCGCTACCTGGGACAGCTTCCCCTTGGGCTACCAAGCTTCCGGAGGCCG 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGTGCAGGCCGCTACCTGGGACAGCTTCCCCTTGGGCTACCAAGCTTCCGGAGGCCG 360

Qy    361 TGCTATTCCTGGGGGTGTGCGCGCCATCTGGGCCCTCGTGTCACCTGGGTCTG 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGCTATTCCTGGGGGTGTGCGCGCCATCTGGGCCCTCGTGTCACCTGGGTCTG 420

Qy    421 GTCTTTGGGTTGGAGGCTCCAGGAGCTCCTGGTCGTCTGCGGCCTCTGCGGGC 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GTCTTTGGGTTGGAGGCTCCAGGAGCTCCTGGTCGTCTGCGGCCTCTGCGGGC 480

Qy    481 AACACACCGGTCAACGGCTTCTCTCTCCGGCACTGCGCCCGCCGACGGCGAAGCTG 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AACACACCGGTCAACGGCTTCTCTCTCCGGCACTGCGCCCGCCGACGGCGAAGCTG 540

Qy    541 CCGGCCCGCTTCAGCCTCTTTTTTCTGCCCTTGGCCATCACAGCCTTC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CCGGCCCGCTTCAGCCTCTTTTTTCTGCCCTTGGCCATCACAGCCTTC 600

Qy    601 TGCTACGTGGGCTGCGGGGGGTGTGCGGCCGGCCCCTGCTTCTGCGGTAGGACCCTAC 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TGCTACGTGGGCTGCGGGGGGTGTGCGGCCGGCCCCTGCTTCTGCGGTAGGACCCTAC 660

Qy    661 CGGGGCCGCTGGGTGCCGCCGGGGGGCCTCCTCACGCTGCTGCGTAGGACCCTAC 720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 CGGGGCCGCTGGGTGCCGCCGGGGGGCCTCCTCACGCTGCTGCGTAGGACCCTAC 720

Qy    721 AACGCCTCCAACGTGGCCAGCTTCCTGTACCCCAATCTAGGAGGCTTCTGGCGAAGCTG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 AACGCCTCCAACGTGGCCAGCTTCCTGTACCCCAATCTAGGAGGCTTCTGGCGAAGCTG 780

Qy    781 GGGCTCATCACGGGTGCCTGGAGTGTGGTGCTTAATCCGCTGGTGACCGGTTACTTGGGA 840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GGGCTCATCACGGGTGCCTGGAGTGTGGTGCTTAATCCGCTGGTGACCGGTTACTTGGGA 840

Qy    841 AGGGGTCCTGGCCTGAAGACAGTGTGCGGCAAGAACGCAAGGGGGCAAGTCCCAGAAG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 AGGGGTCCTGGCCTGAAGACAGTGTGCGGCAAGAACGCAAGGGGGCAAGTCCCAGAAG 900

Qy    901 TAA 903
          |||
Db    901 TAA 903
```

RESULT 2
US-10-225-567A-474
; Sequence 474, Application US/10225567A
; Publication No. US20030113798A1
; GENERAL INFORMATION:
;   APPLICANT: LifeSpan Biosciences
;   APPLICANT: Brown, Joseph P.
;   APPLICANT: Burmer, Glenna C.
;   APPLICANT: Roush, Christine L.
;   TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS
;   FILE REFERENCE: 1920-4-4
;   CURRENT APPLICATION NUMBER: US/10/225,567A
;   CURRENT FILING DATE: 2001-12-19
;   PRIOR APPLICATION NUMBER: 60/257,144
;   PRIOR FILING DATE: 2000-12-19
;   NUMBER OF SEQ ID NOS: 2292
;   SOFTWARE: PatentIn version 3.1

```
Db    121 AGCCTGGTCACGCCCTGAACCTGGCTGCTCCGACCTGCTGCTGCTGACAGTCTCTGCCC 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    121 AGCCTGGTCACGCCCTGAACCTGGCTGCTCCGACCTGCTGCTGCTGACAGTCTCTGCCC 180

Qy    181 CTGAAGGCCGTGGAGGCGCTAGCCTGTGCCCCTGGCTAGCCTTCCGGAGGCCG 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CTGAAGGCCGTACCTGGGACAGCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCG 240

Qy    241 GTCTTCGCGGTGGCCCACTCTTCCCACTCTATGCCGGGGGCTTCTGGCCGCCCCTG 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GTCTTCGCGGTGGCCCACTCTTCCCACTCTATGCCGGGGGCTTCTGGCCGCCCCTG 300

Qy    301 AGTGCAGGCCGCTACCTGGGACAGCTTCCCCTTGGGCTACCAAGCTTCCGGAGGCCG 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGTGCAGGCCGCTACCTGGGACAGCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCG 360

Qy    361 TGCTATTCCTGGGGGTGTGCGCGCCATCTGGGCCCTCGTGTCACCTGGGTCTG 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGCTATTCCTGGGGGTGTGCGCGCCATCTGGGCCCTCGTGTCACCTGGGTCTG 420

Qy    421 GTCTTTGGGTTGGAGGCTCCAGGAGCTCCTGGTCGTCTGCGGCCTCTGCGGGC 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GTCTTTGGGTTGGAGGCTCCAGGAGCTCCTGGTCGTCTGCGGCCTCTGCGGGC 480

Qy    481 AACACACCGGTCAACGGCTTCTCTCTCCGGCACTGCGCCCGCCGACGGCGAAGCTG 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AACACACCGGTCAACGGCTTCTCTCTCCGGCACTGCGCCCGCCGACGGCGAAGCTG 540

Qy    541 CCGGCCCGCTTCAGCCTCTTTTTTCTGCCCTTGGCCATCACAGCCTTC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CCGGCCCGCTTCAGCCTCTTTTTTCTGCCCTTGGCCATCACAGCCTTC 600

Qy    601 TGCTACGTGGGCTGCGGGGGGTGTGCGGCCGGCCCCTGCTTCTGCGGTAGGACCCTAC 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TGCTACGTGGGCTGCGGGGGGTGTGCGGCCGGCCCCTGCTTCTGCGGTAGGACCCTAC 660

Qy    661 CGGGGCCGCTGGGTGCCGCCGGGGGGCCTCCTCACGCTGCTGCGTAGGACCCTAC 720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 CGGGGCCGCTGGGTGCCGCCGGGGGGCCTCCTCACGCTGCTGCGTAGGACCCTAC 720

Qy    721 AACGCCTCCAACGTGGCCAGCTTCCTGTACCCCAATCTAGGAGGCTTCTGGCGAAGCTG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 AACGCCTCCAACGTGGCCAGCTTCCTGTACCCCAATCTAGGAGGCTTCTGGCGAAGCTG 780

Qy    781 GGGCTCATCACGGGTGCCTGGAGTGTGGTGCTTAATCCGCTGGTGACCGGTTACTTGGGA 840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GGGCTCATCACGGGTGCCTGGAGTGTGGTGCTTAATCCGCTGGTGACCGGTTACTTGGGA 840

Qy    841 AGGGGTCCTGGCCTGAAGACAGTGTGCGGCAAGAACGCAAGGGGGCAAGTCCCAGAAG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 AGGGGTCCTGGCCTGAAGACAGTGTGCGGCAAGAACGCAAGGGGGCAAGTCCCAGAAG 900

Qy    901 TAA 903
          |||
Db    901 TAA 903
```

RESULT 3
US-10-029-386-22904
; Sequence 22904, Application US/10029386
; Publication No. US2003194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 22904
; LENGTH: 903
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO U62631.1
; OTHER INFORMATION: EXPRSSED IN HELA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.89
; OTHER INFORMATION: NT HIT: gi11526068, EVALUE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: O14842, EVALUE 0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: AW583167.1, EVALUE 0.00e+00
US-10-029-386-22904

Query Match 100.0%; Score 903; DB 15; Length 903;
Best Local Similarity 100.0%; Pred. No. 3.8e-210;
Matches 903; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 ATGGACCTGCCCCCGCAGCTCTCTTCGGCCTCTATGTGCCGCCGTTGCGCTGGGCTTC 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGGACCTGCCCCCGCAGCTCTCTTCGGCCTCTATGTGCCGCCGTTGCGCTGGGCTTC 60

Qy 61 CCGCTCAACGTCCTGGCCATCGAGGCGCGACGGCCCAGCCCGGCTCCGTCTCACCCCT 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 61 CCGCTCAACGTCCTGGCCATCGAGGCGCGACGGCCCAGCCCGGCTCCGTCTCACCCCT 120

Qy 121 AGCCTGGTCTACGCCCTGAACCTGGGCTGCTCCGACCTGCTGACAGTCTCTCTGCCC 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 AGCCTGGTCTACGCCCTGAACCTGGGCTGCTCCGACCTGCTGACAGTCTCTCTGCCC 180

Qy 181 CTGAAGGCCGGTGGAGGGCGCTAGCCTCCGGGGCCTGGCCTCTGCCGGCCGTGCCCC 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 CTGAAGGCCGGTGGAGGGCGCTAGCCTCCGGGGCCTGGCCTCTGCCGGCCGTGCCCC 240

Qy 241 GTCTTCGCGGTGGCCCCACTTCTTCCCACTCTTCCCACTCTATGTGCCGGCGGGGCTTC 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GTCTTCGCGGTGGCCCCACTTCTTCCCACTCTATGTGCCGGCGGGGGCTTC 300

Qy 301 AGTGCAGGCCGCTACCTGGGAGCAGCCTTCCCCTTGGGCTACAAAGCCTTCCGGAGGCCG 360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 AGTGCAGGCCGCTACCTGGGAGCAGCCTTCCCCTTGGGCTACAAAGCCTTCCGGAGGCCG 360

Qy 361 TGCTATTCCTGGGGGGTGTGCGCGGCCATCGGGCCCCGTCTGTGTCACCTGGGTCTG 420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 TGCTATTCCTGGGGGGTGTGCGCGGCCATCGGGCCCCGTCTGTGTCACCTGGGTCTG 420

Qy 421 GTCTTTGGGTTGGAGGCTCCAGGAGGCTGGCTGACCCACAGCAACACCTCCCTGGGCATC 480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421 GTCTTTGGGTTGGAGGCTCCAGGAGGCTGGCTGACCCACAGCAACACCTCCCTGGGCATC 480

Qy 481 AACACACCGGTCAACGGCTCTCCGGTCTCCGGAGGCCTGGGACCCGGCCTCTGCCGGC 540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 481 AACACACCGGTCAACGGCTCTCCGGTCTCCGGAGGCCTGGGACCCGGCCTCTGCCGGC 540

Qy 541 CCGGCCCGCTTCAGCCTCTTCTCCTGCTCTTTTTTCTGCCCTTGGCCATCACAGCCTTC 600
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 541 CCGGCCCGCTTCAGCCTCTTCTCCTGCTCTTTTTTCTGCCCTTGGCCATCACAGCCTTC 600

Qy 601 TGCTACGTGGCTGCTCCGGCCACTGGCCGCTCCGGCCTCTGACGCACAGGCGGAAGCTG 660
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 601 TGCTACGTGGCTGCTCCGGCCACTGGCCGCTCCGGCCTCTGACGCACAGGCGGAAGCTG 660

Qy 661 CGGGGCCGCCGCCTGGGTGGCCGGCGGGGGGCCCTCCTCCAGCGTGCTGCTCCGTAGGACCCTAC 720
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 661 CGGGGCCGCCGCCTGGGTGGCCGGCGGGGGGCCCTCCTCCAGCGTGCTGCTCCGTAGGACCCTAC 720

Qy 721 AACGCCTCCAACGTGCCAGCTTCGTACCCCAATCTAGGAGGCTCCTGGCGGAAGCTG 780
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 721 AACGCCTCCAACGTGCCAGCTTCCTGTACCCCAATCTAGGAGGCTCCTGGCGGAAGCTG 780

Qy 781 GGGCTCATCACGGGTGCCTGGAGTGGTGGTGCTTAATCCGCTGGTGACCGGTTACTGGGA 840
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 781 GGGCTCATCACGGGTGCCTGGAGTGGTGGTGCTTAATCCGCTGGTGACCGGTTACTTGGGA 840

Qy 841 AGGGGTCCTGGCCTGAAGACAGTGTGTGCGGCAAGAACGCAAGGGGGCAAGTCCCAGAAG 900
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 841 AGGGGTCCTGGCCTGAAGACAGTGTGTGCGGCAAGAACGCAAGGGGGCAAGTCCCAGAAG 900

Qy 901 TAA 903
         |||
Db 901 TAA 903

RESULT 4
US-10-202-687-1
; Sequence 1, Application US/10202687
; Publication No. US20040019109A1
; GENERAL INFORMATION:
; APPLICANT: OWMAN, CHRISTER
; APPLICANT: OLDE, BJORN
; APPLICANT: KOTARSKY, KNUT
; APPLICANT: NILSSON, NICLAS
; APPLICANT: FLODGREN, ERIK
; TITLE OF INVENTION: METHODS OF IDENTIFYING COMPOUNDS AFFECTING FATTY ACID
; TITLE OF INVENTION: METABOLISM
; FILE REFERENCE: 07675.0007 SEQUENCE LISTING
; CURRENT APPLICATION NUMBER: US/10/202,687
; CURRENT FILING DATE: 2002-07-24
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 903
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-202-687-1

Query Match 100.0%; Score 903; DB 16; Length 903;
Best Local Similarity 100.0%; Pred. No. 3.8e-210;
Matches 903; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 ATGGACCTGCCCCCGCAGCTCTCTTCGGCCTCTATGTGCCGCCGTTGCGCTGGGCTTC 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGGACCTGCCCCCGCAGCTCTCTTCGGCCTCTATGTGCCGCCGTTGCGCTGGGCTTC 60

Qy 61 CCGCTCAACGTCCTGGCCATCCGAGGCGCGAGCGGCCCAGCCCGGCTCCGTCTCGTTCACCCCCT 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 61 CCGCTCAACGTCCTGGCCATCCGAGGCGCGAGCGGCCCAGCCCGGCTCCGTCTCCACCCCT 120

Qy 121 AGCCTGGTCTACGCCCTGAACCTGGGCTGCTCCGACCTGCTGTCTGACAGTCTCTCGCCC 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 AGCCTGGTCTACGCCCTGAACCTGGGCTGCTCCGACCTGCTGTCGACAGTCTCTCGCCC 180

Qy 181 CTGAAGGCCGGTGGAGGGCGCTAGCCTCCGGGGCCTCCGGGGGGCTTCCGGCGGGGCCCCGCCCTG 240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 CTGAAGGCCGGTGGAGGGCGCTAGCCTCCGGGGCCTCCGGGGGGCTTCCGGCGGGGCCCCGCCCTG 240

Qy 241 GTCTTCGCGGTGGCCCCACTTCTTCCACTCTTATGCCGGGGGGGCTTCCTGGCCGCCCTG 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GTCTTCGCGGTGGCCCCACTTCTTCCACTCTTATGCCGGGGGGGGCTTCCTGGCCGCCCTG 300

```
Qy  301 AGTGCAGGCCGCTACCTGGGAGCAGCCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCG 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 AGTGCAGGCCGCTACCTGGGAGCAGCCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCG 360

Qy  361 TGCTATTCCTCGGGGGTGTGCGCGGCCCATCTGGGCCCTCGTCCTGTGTCACCTGGGTCTG 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 TGCTATTCCTCGGGGGTGTGCGCGGCCCATCTGGGCCCTCGTCCTGTGTCACCTGGGTCTG 420

Qy  421 GTCTTTGGGTTGGAGGCTCCAGAGAGGCTGGCTGGACCACAGCAACACACTCCCTGGGCATC 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 GTCTTTGGGTTGGAGGCTCCAGAGAGGCTGGCTGGACCACAGCAACACACTCCCTGGGCATC 480

Qy  481 AACACACCGGTCAACGGCTCTCCGTCTGCCTGGAGGCCTGGGACCCGGCCTCTGCCCGC 540
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 AACACACCGGTCAACGGCTCTCCGTCTGCCTGGAGGCCTGGGACCCGGCCTCTGCCCGC 540

Qy  541 CCGGCCCGCTTCAGCCTCTGCTCTTTTTCTGCCCCTTGGCCATCACAGCCTTC 600
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 CCGGCCCGCTTCAGCCTCTGCTCTTTTTCTGCCCCTTGGCCATCACAGCCTTC 600

Qy  601 TGCTACGTGGGCTGGCTGCCTCCGGGCACTGGGCCCGCTCCGGGCCTGAGCGCAGGCGAAGCTG 660
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 TGCTACGTGGGCTGGCTGCCTCCGGGCACTGGGCCCGCTCCGGGCCTGAGCGCAGGCGAAGCTG 660

Qy  661 CGGGCCCGCGTTGGCCGGCGGCCCTCCTCACGCGCTCGGTAGGACCCTAC 720
        |||||||||||||||||||||||||||||||||||||||||||||||
Db  661 CGGGCCCGCGTTGGCCGGCGGCCCTCCTCACGCGCTCGGTAGGACCCTAC 720

Qy  721 AACGGCTCCAACGTGGCCAGCTTCCGTACCCCAATCTAGGAGGCTCCTGCGGAAGCTG 780
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 AACGGCTCCAACGTGGCCAGCTTCCGTACCCCAATCTAGGAGGCTCCTGCGGAAGCTG 780

Qy  781 GGGCTCATCACGGGTGCCTGGAGTGTGTGCTTAATCCGGTGGTGACCGGTTACTTGGGA 840
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 GGGCTCATCACGGGTGCCTGGAGTGTGTGCTTAATCCGGTGGTGACCGGTTACTTGGGA 840

Qy  841 AGGGGTCCTGGCCTGAAGACAGTGTGCGGCAAGAACGCAAGGGGGCAAGTCCCAGAAG 900
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  841 AGGGGTCCTGGCCTGAAGACAGTGTGCGGCAAGAACGCAAGGGGGCAAGTCCCAGAAG 900

Qy  901 TAA 903
        |||
Db  901 TAA 903
```

```
Qy    1 ATGGACCTGCCCCGCAGCTCTCCTTCGGCTCTTATGTGGCCGGCCTTGCGGCTGGGCTTC 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGGACCTGCCCCGCAGCTCTCCTTCGGCTCTTATGTGGCCGGCCTTGCGGCTGGGCTTC 60

Qy   61 CCGCTCAACGTCTCTGGCCATCCGAGGCGGCGACGGCCCACGCCGGCCTCCGTCCGCCCT 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CCGCTCAACGTCTCTGGCCATCCGAGGCGGCGACGGCCCACGCCGGCCTCCGTCCGCCCT 120

Qy  121 AGCCTGGTCTACGCCCTGAACCTGGGCTGCTGCGACCTGGCTGCTGACAGTCTCTCGCCC 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 AGCCTGGTCTACGCCCTGAACCTGGGCTGCTGCGACCTGGCTGCTGACAGTCTCTCGCCC 180

Qy  181 CTGAAGGCGGTGGAGGCGCTAGCCTTCGGGGCCCTCGGCCCTCGCGCCCTCGTGCCCC 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 CTGAAGGCGGTGGAGGCGCTAGCCTTCGGGGCCCTCGGCCCTCGCGCCCTCGTGCCCC 240

Qy  241 GTCTTCGCCGGTGGCGCCCACTTCTTCCCACACTCTATGCCGGCGGCGGGGCCTCCGGCCG 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 GTCTTCGCCGGTGGCGCCCACTTCTTCCCACACTCTATGCCGGCGGCGGGGCCTCCGGCCG 300

Qy  301 AGTGCAGGCCGCTACCTGGGAGCAGCCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCG 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 AGTGCAGGCCGCTACCTGGGAGCAGCCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCG 360

Qy  361 TGCTATTCCTCGGGGGTGTGCGCGGCCCATCTGGGCCCTCGTCCTGTGTCACCTGGGTCTG 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 TGCTATTCCTCGGGGGTGTGCGCGGCCCATCTGGGCCCTCGTCCTGTGTCACCTGGGTCTG 420

Qy  421 GTCTTTGGGTTGGAGGCTCCAGAGAGGCTGGCTGGACCACAGCAACACACTCCCTGGGCATC 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 GTCTTTGGGTTGGAGGCTCCAGAGAGGCTGGCTGGACCACAGCAACACACTCCCTGGGCATC 480

Qy  481 AACACACCGGTCAACGGCTCTCCGTCTGCCTGGAGGCCTGGGACCCGGCCTCTGCCCGGC 540
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 AACACACCGGTCAACGGCTCTCCGTCTGCCTGGAGGCCTGGGACCCGGCCTCTGCCCGGC 540

Qy  541 CCGGCCCGCTTCAGCCTCTGCTCTTTTTCTGCCCCTTGGCCATCACAGCCTTC 600
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 CCGGCCCGCTTCAGCCTCTGCTCTTTTTCTGCCCCTTGGCCATCACAGCCTTC 600

Qy  601 TGCTACGTGGGCTGGCTGCCTCCGGGCACTGGGCCCGCTCCGGGCCTGAGCGCAGGCGAAGCTG 660
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 TGCTACGTGGGCTGGCTGCCTCCGGGCACTGGGCCCGCTCCGGGCCTGAGCGCAGGCGAAGCTG 660

Qy  661 CGGGCCCGCGTTGGCCGGCGGCCCTCCTCACGCGCTCGGTAGGACCCTAC 720
        |||||||||  ||||||||||||||||||||||||||||||||||||||
Db  661 CGGGCCAAATGGGTGGCCGGCGGCCCTCCTCACGCGCTCGGTAGGACCCTAC 720

Qy  721 AACGGCTCCAACGTGGCCAGCTTCCGTACCCCAATCTAGGAGGCTCCTGCGGAAGCTG 780
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 AACGGCTCCAACGTGGCCAGCTTCCGTACCCCAATCTAGGAGGCTCCTGCGGAAGCTG 780

Qy  781 GGGCTCATCACGGGTGCCTGGAGTGTGTGCTTAATCCGGTGGTGACCGGTTACTTGGGA 840
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 GGGCTCATCACGGGTGCCTGGAGTGTGTGCTTAATCCGGTGGTGACCGGTTACTTGGGA 840

Qy  841 AGGGGTCCTGGCCTGAAGACAGTGTGCGGCAAGAACGCAAGGGGGCAAGTCCCAGAAG 900
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  841 AGGGGTCCTGGCCTGAAGACAGTGTGCGGCAAGAACGCAAGGGGGCAAGTCCCAGAAG 900

Qy  901 TAA 903
        |||
Db  901 TAA 903
```

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 9201
; LENGTH: 550
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO U62631.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.89
; OTHER INFORMATION: SWISSPROT HIT: O14842, EVALUE 1.00e-22
; OTHER INFORMATION: EST_HUMAN HIT: AU117321.1, EVALUE 2.00e-01
; OTHER INFORMATION: NT HIT: U62631.1, EVALUE 0.00e+00
US-10-029-386-9201

Query Match   17.1%;   Score 154; DB 15; Length 550;
Best Local Similarity 100.0%;   Pred. No. 3.9e-28;
Matches   154; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 ATGGACCTGCCCCCCGCAGCTCTCCTTCGGCCTCTATGTGGCCGCCTTGCGCTGGCTTC   60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   397 ATGGACCTGCCCCCCGCAGCTCTCCTTCGGCCTCTATGTGGCCGCCTTGCGCTGGCTTC 456

Qy   61 CCGGCTCAACGTCCTGGCCATCCGAGCGCGACAGGCGCCACGCCCCGGCTCCGTCTCACCCCT 120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   457 CCGGCTCAACGTCCTGGCCATCCGAGCGCGACAGGCGCCACGCCCCGGCTCCGTCTCACCCCT 516

Qy   121 AGCCTGGTCTACGCCCTGAACCTGGGCTGCTCCG 154
            ||||||||||||||||||||||||||||||||||
Db   517 AGCCTGGTCTACGCCCTGAACCTGGGCTGCTCCG 550

RESULT 7
US-10-029-386-24088
; Sequence 24088, Application US/10029386
; Publication No. US2003019470A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 24088
; LENGTH: 960
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC002511.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.5
; OTHER INFORMATION: SWISSPROT HIT: O15529, EVALUE 0.00e+00
; OTHER INFORMATION: NT HIT: AF024689.1, EVALUE 0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: AA860692.1, EVALUE 2.20e-02
US-10-029-386-24088

Query Match   13.5%;   Score 121.8; DB 15; Length 960;
Best Local Similarity 52.1%;   Pred. No. 2.5e-20;
Matches   382; Conservative   0; Mismatches 327; Indels 24; Gaps   4;

Qy   18 GCTCTCCTTCGGCCTCTATGTGGCCGCCTTGCGCTGGCTTCCCGCTCAACGTCCTGGC   77
            ||||| ||||| | ||||||| | | || |||| ||||||||| |||| ||||
Db   56 GTTCGCTCTTCTCGGTGTACCTTCTCCACTTTCCTCGGTGTGGGGGCTCCCCTCAACCTGTGGC 115

; OTHER INFORMATION:
; OTHER INFORMATION:
; OTHER INFORMATION:
; FEATURE:
; ORGANISM: Homo sapiens
; TYPE: DNA
; LENGTH: 1041
; SEQ ID NO 253
; SOFTWARE: Patentin version 3.1
; NUMBER OF SEQ ID NOS: 294
; PRIOR APPLICATION NUMBER: US/09/170,496
; PRIOR FILING DATE: 1998-10-13
; CURRENT FILING DATE: 2002-09-20
; CURRENT APPLICATION NUMBER: US/10/251,385
; FILE REFERENCE: AREN-0040
; TITLE OF INVENTION: Receptors
; TITLE OF INVENTION: Protein-Coupled
; TITLE OF INVENTION: US20030105292A1-Endogenous, Constitutively Activated Human G
; APPLICANT: Liaw, Chen W.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Behan, Dominic P.
; GENERAL INFORMATION:
; Publication No. US20030105292A1
; Sequence 253, Application US/10251385
US-10-251-385-253
RESULT 8

Qy   78 CATCCGAGGCGGACGGCCCACGCCCCGGCTCCGTCTCACCCGTCGTCGTCTTACCGCCCT 137
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   116 CCTGGTGGTCTTCGTGGGCAAGCTGCGGTGCCGCCCGGTGCCGTGGACGTGCTCCTGCT 175

Qy   138 GAACCTGGCTGCTCCGACGACCTGCTCCTGCTGGACAGTCTCTGTCCTGCTGTTCCGTGGAGGC 197
            ||||||| |||||||||||||| ||| |||||||| |||||||||||||||||
Db   176 CAACCTGACCCGCTCCGGACCTGTCCTGCTGGCCCTTTCCGCATGGTGGAGGC 235

Qy   198 GCTAGCCTCCGGGGCCTGGGCTTCGGACCTCGTGCCCGCTTCGCGGTGGCGGCCCA 257
            |||| ||| ||||||| ||||| ||||||| ||||| || || |||
Db   236 AGCCAATGGCATGCACTGCGCCCGGCCCCTTCATCCTCTGCCCACTCTCTGGATTCATCTT 295

Qy   258 CTTCTTCCACTCTATGCCGGGGGCTTCGGCGCCCCTGAGTGCAGGCCGCTACCT 317
            |||||||| | |||||| ||||| |||||||||||| |||| ||||||||||
Db   296 CTTCACCACCATCTATCTCACCGGCCCTTCCTGGCAGCTGTGAGCATTGAACGCTTCCT 355

Qy   318 GGGAGCAGCCTTCCCCTGGGCTACCAAGCCTCCGAGGCCGTGTATTCCTGGGGGT 377
            |||| |||||| | ||||| ||| |||| || |||| ||| | ||||||||
Db   356 GAGTGTGGGCCCATCTGGCCCCTCGTCTCGTGTCACCTGGGTCGGTTGGGTTGGAGGC 415

Qy   378 GTGCGCGGCCATCTGGGCCCTCGTCTCGTGTCACCTGGGTCGGTTGGGTTGGAGGC 437
            |||| |||| || |||||| ||||||||||||||| ||||||||||||
Db   416 GAGTGTGGCCTGCCTGTGTTGGCCTCTGCACTGCAGCTGGTCTCACTGACGTGGTCATAGAATT 475

Qy   438 TCCAGGAGGCTGGCTGACCACAGCAACACTCCCTGGGCATCAACAACACCGGTCAACGG 497
            |||||||||||| || ||||||| || ||||||||||
Db   476 ---------ACATCTCCCACAGCCAGGCACCAATGGGACCTG 517

Qy   498 CTCTCCGGTCTGCCTGGAGGCCTGGACCGGTGGAGGCCGCCGCCTTCAGCCT 557
            || ||||||| |||| |||| |||||| ||| | |||||| |||||||||
Db   518 CTACCTGGAGTTCTGGAAGGACCAG--CTAGCCATCCTCCTGCCCGTGCGGGCTGGAGA-T 574

Qy   558 CTCTCCGTCTTTTTCTGCCCTTGGCCATCACAGCCTTCTGCTACGTGGGCTGCCT 617
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   575 GGCTGTGGTCCTCTTTTGTGGTCCCGCGATCATCACCAGCTACTGCTACACAGCGCCGTGGT 634

Qy   618 CCGGGCACTGCCCGCTCCGGCTGACGCACAGGCCGGAAGCTGCGGCGCCTGGGTGGC 677
            ||||| ||||| | ||||| ||||||||||| |||| |||| ||||||||||||
Db   635 GTGGATCCTCCGGCAGAGGGGGC---AGCCACCGCCGGCGCAGAGGAGGGTGGCGGGGGCTGGT 691

Qy   678 CGGCGGGGCCCTCCCACGCGTGGCTCAACTTCCTGTCTGTTGGGCCCTACAACGTGCTGGC 737
            ||||||||||||||||||||| |||||||||||||||||| ||||||||||||||||
Db   692 GGCGGCCACGCTGCTCAACTTCCTGTCTGTTGGGCCCTACAACGTGTCCCATGTCGT 751

Qy   738 CAGCTTCCTGTAC 750
            |||| |||||||
Db   752 GGGCTATATCTGC 764

```
Query Match      13.5%;   Score 121.8;   DB 15;   Length 1041;
Best Local Similarity   52.1%;   Pred. No. 2.5e-20;
Matches 382;   Conservative   0;   Mismatches 327;   Indels  24;   Gaps  4;

Qy   18  GCTCTCCTTCGGCCTCTATGTGGCCGCCTTTGCGCTGGGCTTCCCGCTCAACGTCCTGGC   77
             ||| ||       ||   | |  ||||  | | ||| || |||| |||||   || |||
Db   45  GTTCGTCTTCTCGGTGTACCTTCTCACTTTCTCGTGTGGGCTCCCCTCAACCTGCTGGC  104

Qy   78  CATCCGAGGCGGCGACGGCCCACGCCCGGCTCCGTCTCACCCTGCTGTCTACGCCCT  137
             ||  || || | | |||| || | |||  |  ||  || || || |  | |||| |
Db  105  CCTGGTGGTCTTCGTGGGCAAGCTGCAGCCTGCGCCGCCGGTGCGCCTCTCGT  164

Qy  138  GAACCTGGGCTGCTCCGACCTGCTGCTGACAGTCTCTCTGCCCCTGAAGGCGGTGGAGGC  197
             || ||||  ||  || || || ||| |  | | |||  |||| | |||| | |||||
Db  165  CAACCTGACCGCCTCGGACCTGCTCCTGCGTGTTCCTGCCTTTCCGCTGTGTGGGAGGC  224

Qy  198  GCTAGCCTCCGGGGCCTGGCCTCTGCCGCCTGAGTGCAGGCCGCTACCT  257
             || ||| | || |||| ||||| | |  | ||||| |  | ||  || |
Db  225  AGCCAATGGCATGCACTGGCCCTCCATCCTCCGCCCACTCTCTCGGATTCATCTT  284

Qy  258  CTTCTTCCCACTCTATGCCGGCGGGGGCCTTCCTGGAGGCCGTGCTATTCCTGGGGGGT  317
             |||  |||  || ||| ||| | ||||| |||| ||| ||   |||  || ||| |||
Db  285  CTTCACCACCATCTATCTCACCGCCCTCTTCCTGGCTGCTGGCATTGAACGCTTCCT  344

Qy  318  GGGAGCAGCCATCTGGGCCTGGACCACAGCAACACTCCCTGGGGT  377
             ||| |||| || ||| || ||| | || |||| ||||| ||
Db  345  GAGTGTGGCCACCACTGTGGTACAAGACCCGGCGGGCAGGTCTGGT  404

Qy  378  GTGCGCGGCCATCTGGGCCTGGACCACGGTCTTTGGGTTGGAGGC  437
             || || ||| ||| ||| |||  | || | ||| ||| || |||
Db  405  GAGTGTGGCCTGTGTTGGCGTGGTTGGCCGTCACTGCAGCGTGGGTGCTGGT  464

Qy  438  TCCAGGAGGCTGGCTGGACCACAGCAACACTCCCTGGGCATCAACACACGG  497
             |||   |   || || | | |      ||| |||| || |||||  ||||
Db  465  CTCAGGGG-------------ACATCTCCACCAGCCAGGGCACCAATGGGACCTG  506

Qy  498  CTCTCCGGTCTCTTTTTTCTGCGCCTGCCATCACAGCCTTCTGCTACGTGGGCTGCCT  557
             |||||| | | ||||||| ||| ||||| |   |||| | |||  |||| || |  ||
Db  507  CTACCTGGAGTTCCGGAAGGACCAG--CTAGCCATCCTCCTGCCCGGCTGAG-AT  563

Qy  558  CTCTCCTGCTCTTTTTCGCCCTCTGCGTAGGACCCTCACAACGCCTCCAACGTGC  617
             ||||| | | ||   |  | ||  ||| | |  ||| |    | | | | |  ||
Db  564  GGCTGGTCGGCCTGGTTTGTGTCCCGCGATCATCACCAGCTACTGCTAC---AGCCGCCT  620

Qy  618  CCGGGCACTGGCCCGCTCCACGGCGTGCTGCTGCGTAGGACCCTCCAAACGTGGC  677
             || ||||  |  || | || | | | ||| || || || ||  || | | ||
Db  621  GGTGTGGATCCTCGGCAGAGAGGGCAGCCACCGCCGCCTAGGAGCCGTGGGGGCTGTT  680

Qy  678  GGCGGCCACGCTGCTCAACTTCCTTGTCTGTTTGGGCCCTACAACGTGTCCCATGTCGT  737
             | | ||||||  | || |  |  |  ||| ||| |||  | |||| |||| ||| |||
Db  681  CGGCGGCCACGCGTCCTCCACCGCTCTGCGTAGGACACCGCTGCGTGGCCCTTCCAACGTCGT  740

Qy  738  CAGCTTCCTGTAC  750
             |||  | || |||
Db  741  GGGCTATATCTGC  753
```

```
; NUMBER OF SEQ ID NOS: 2292
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 604
; LENGTH: 1041
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-225-567A-604

Query Match      13.5%;   Score 121.8;   DB 15;   Length 1041;
Best Local Similarity   52.1%;   Pred. No. 2.5e-20;
Matches 382;   Conservative   0;   Mismatches 327;   Indels  24;   Gaps  4;
```

RESULT 9
US-10-225-567A-604
; Sequence 604, Application US/10225567A
; Publication No. US20030113798A1
; GENERAL INFORMATION:
; APPLICANT: LifeSpan Biosciences
; APPLICANT: Brown, Joseph P.
; APPLICANT: Burmer, Glenna C.
; APPLICANT: Roush, Christine L.
; TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 1920-4-4
; CURRENT APPLICATION NUMBER: US/10/225,567A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/257,144
; PRIOR APPLICATION FILING DATE: 2000-12-19

RESULT 10
US-10-203-539-1
; Sequence 1, Application US/10203539
; Publication No. US20030113810A1
; GENERAL INFORMATION:
; APPLICANT: GLAXO GROUP LIMITED
; TITLE OF INVENTION: NOVEL ASSAY
; FILE REFERENCE: PG3849USw

RESULT 11
US-10-203-539-3
; Sequence 3, Application US/10203539
; Publication No. US20030113810A1
; GENERAL INFORMATION:
; APPLICANT: GLAXO GROUP LIMITED
; TITLE OF INVENTION: NOVEL ASSAY
; FILE REFERENCE: PG3849USw
; CURRENT APPLICATION NUMBER: US/10/203,539
; CURRENT FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: GB 0003900.8
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: GB 0007015.1
; PRIOR FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 3.0
;
; SEQ ID NO 3
; LENGTH: 1041
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1041)
US-10-203-539-3

Query Match   13.5%;   Score 121.8; DB 15;   Length 1041;
Best Local Similarity 52.1%;   Pred. No. 2.5e-20;
Matches 382; Conservative 0; Mismatches 327; Indels 24; Gaps 4;

Qy   18 GCTCTCCTTCGGCCTCTATGTGGCCGCCTTTGCGCTGGGCTTCCCGCTCAACGTCCTGGC 77
         || |||| ||| |||||| | || || || ||| || ||||||| |||||||
Db   45 GTTCGTCTTCTCGGTGTACCTTTCTCACTTTCCTGGTGTGGGGCTCCCCCTCAACCTGCTGGC 104

Qy   78 CATCCGAGGCGGCGACGGCCCAGCCCGGCTCCGTCTCACCCCTAGCCTGGTCTACGCCCT 137
         ||||| ||| | |||| || || ||| || | ||||||||||||||||||| ||
Db   105 CCTGGTGGGTCTTCGTGGGCAAGCTGCAGCGCGCCGGTGGCCGTGGACGTGCTCCTGCT 164

Qy   138 GAACCTGGCTGCTCCGACCAGTCTCTCTGCCCTGAAGGCGGTGGAGGC 197
         ||| |||| | ||||| || | | ||| || ||||||||||
Db   165 CAACCTGACCGCCTCGGACCTGCTCCTGCTGTTCCTGCCTTTCCGCATGGTGGAGGC 224

Qy   198 GCTAGCCTCCGGGGCCTGGCCTTCGCGTCTTCGCGGCTCGGTCGGTGGGGTCGGCCA 257
         ||||||| | | | |||| || |||| | ||||| ||||||||||| |||||
Db   225 AGCCAATGGCATGCACCTGGTACAAGACCCGGCCGAGGCGTGGGGCTCTGGT 284

Qy   258 CTTCTTCCCACTCTATGCCGGCGGGGGTTCCTGGCCGCCGCCCTGAGTGCAGGCCGCTACCT 317
         ||||||||| | | | || || || ||||| || ||| | | ||||| |||||||
Db   285 CTTCACCACCATCTATCTCACGCGCCCTTCCTGGCAGCTGTGTGAGCATTGAACGCTTCCT 344

Qy   318 GGGAGCAGCCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCGGTGCTATTCCTGGGGGGT 377
         |||| ||| | || ||| | | | ||| || |||| || |||| || |||||||
Db   345 GAGTGTGGCCCACCACTGTGGTACAAGACCCGGCCGAGGCGTGGGGCTCTGGT 404

Qy   378 GTGCGCGGCCATCTGGGCCCTGTGTCACCTGGGTCTTTGGGTTGGAGGC 437
         ||| ||| ||||| ||||| | | || | | | |||| ||||||
Db   405 GAGTGTGGCCTGCTGTGGCCTCTGCTGCAGCGGTGTCTCACTGCAGCGTGCTGCTTACGTCATAGAATT 464

Qy   438 TCCAGGAGGCTGGCTGGACCACAGCAACACCTCCCTGGGCATCAACACCGGTCAACGG 497
         |||||| | | ||||||| | ||| ||||||||||||||||||||||
Db   465 CTCAGGG----------------ACATCCCACAGCCAGGGACACCAATGGGACCTG 506

Qy   498 CTCTCCGGTCGCTGCCGGAGGCCTGGACGCACAGCGGCGGAAGTGCGGGTGGTGG 557
         ||||||||||| ||| ||| ||| ||| | ||| || ||||||||||||||
Db   507 CTACCTGGAGTTCTGGAAGGACCAG--CTAGCCATCCTCCTGCGCCCGTGGCGTGGCTGGAGA-T 563

Qy   558 CTCTCTCCTGCTCTTTTTCTGCCCTTGGCCATCACAGCCTTCTCGTACGTGGGCTGCCT 617
         || || | | || | |||||| | | ||||||| || | || |||||
Db   564 GGCTGTGTCCCAGCTGCTGCTTCCTTGTGGTCCCGTGATCATCACCAGCTACTGTCAACGTGTCCATGTCGT 623

Qy   618 CCGGGCACTGGCCCGCTCCGGCCTGACGACGGAAGCTGCGGGGTGCGGC 677
         |||||||| ||| || ||| | ||||| | || || |||| ||||
Db   624 GTGGATCCTCGGCAGAGGGGC---AGCCACCGCCCGGCAGAGAGGGGTGGGGGGGCGGGT 680

Qy   738 CAGCTTCCTGTAC 750
         || ||| | ||
Db   741 GGGCTATATCTGC 753

```
Qy   678  CGGCGGGGGCCCTCCTCAGCTGTCTGCTCTGCGTAGGACCCTACAAACGGCTTCCAACGTGGC  737
          ||||| |||||||||||||||| || | ||| || ||| ||||||||||||||||||||||||
Db   681  GGCGGGCCACGCTGCTCAACTTCCTTGTCTGCTTTGGGCCCTACACACGTGTCCCATGTCGT  740

Qy   738  CAGCTTCCTGTAC  750
          |||  |||||||
Db   741  GGGCTATATCTGC  753
```

RESULT 12
US-10-029-386-25146
; Sequence 25146, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
;   APPLICANT: Penn, Sharron G.
;   APPLICANT: Rank, David R.
;   APPLICANT: Hanzel, David K.
;   TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
;   TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
;   FILE REFERENCE: AEOMICA-X-2
;   CURRENT APPLICATION NUMBER: US/10/029,386
;   CURRENT FILING DATE: 2001-12-20
;   NUMBER OF SEQ ID NOS: 34288
;   SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 25146
;   LENGTH: 1206
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO U62631.1
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.4
;   OTHER INFORMATION: EST HUMAN HIT: BF115946.1, EVALUE 2.80e-02
;   OTHER INFORMATION: SWISSPROT HIT: O14843, EVALUE 0.00e+00
;   OTHER INFORMATION: NT HIT: U62631.1, EVALUE 0.00e+00
US-10-029-386-25146

Query Match   13.5%; Score 121.8; DB 15; Length 1206;
Best Local Similarity   52.1%; Pred. No. 2.4e-20;
Matches 382; Conservative   0; Mismatches 327; Indels   24; Gaps   4;

Qy   18 GCTCTCCTTCGGCCTCTATGTGGCCGCCCTTTGCGCTCCCTTGGCTTCCCGCTCAACGTCCTGC 77
          |||||| || |||||| ||||||| ||| |||| | |||| |||||| |||||
Db   210 GTTCGTCTTCTCGGTGTACCTTCTCACTTTCCTGTGGGGCTCCCCCTCAACCTGCTGC 269

Qy   78 CATCCGAGGCGGACGGGCCCAGCGCCCGGCTCCACCCTAGCTCCTGTCTACGCCCT 137
          |||| |||| | || | ||| |||||||||||| ||||||||||||| |||||
Db   270 CCTGGTGGTCTTCGTGGGCAAGCTGCAGCGGCCGGCTCCACCCTAGCTCCTGTGCTGCTCCTGCT 329

Qy   138 GAACCTGGGCTGCTCCGACCTGTGCTGCTGACCAGTCTCTGCTCCGGAAGGCGGTGGAGGC 197
          |||||||||||| ||||| ||||||||| ||| || | || ||| ||||| |||||||||
Db   330 CAACCTGACCGGCCTCGGACCTGTGCGTCCTGTCCTTTCCGCATGGTGGAGGC 389

Qy   198 GCTAGCCTCCGGGGCCTGGGCCCTGGGGGGCTTCGCGTCTTCGCGTCTTCGGTGGTGGGCCCA 257
          ||||| ||||| |||||| |||| ||| ||| | || ||| ||| |||||||
Db   390 AGCCAATGGCATGCACTGGCCCTGCCCACTCTGCCCACTCTCTGGATTCATCTT 449

Qy   258 CTTCTTCCCACTCTATGCCGCCGCCCTGAGTGAGGACCGCTACCT 317
          ||| | || ||| | ||||| |||||||||| |||||||||||
Db   450 CTTCACCACCATCTATCTCACCGCCCTGCAGCTGTGAGCATTGAACGCTTCCT 509

Qy   318 GGGAGCAGCCTTCCCCCTTGGGCTACCAAGCCTTGCGGAGGCCGTGCTATTTCCTGGGGGGT 377
          |||||| ||||||| |||| |||||||||| ||| ||||||||
Db   510 GAGTGTGGCCCACCCACTGGTGGTCTGTGGTTGGAGCAGCAGGTCTGGT 569

Qy   378 GTGGCGGCGGCAATCTGGGCCTCGTCCGTCACCTGGGTCTTTGGGTTGGAGGC 437
          || ||| |||| ||||| ||| || |||| ||| |||||||| || |||||
Db   570 GAGTGTGGCGCTGCTGCTCTGGCTCACCTGCTCTGGTGTCTCACGTGCTCATAGAATT 629

Qy   438 TCCAGGAGGCTGGCTGGACCACCAGACAACACCTCCTGGCTACCTCCCTGGGCATCAACACCGGTCAACGG 497
          |||| |||||| || | |||
Db   630 CTCAGGGG-------------ACATCTCCCAACAGCCAGGGCACCCAGGACCTG 671
```

```
Qy   498  CTCTCCGGTCTGCTGCTGGAGGCCTGGACCCGGCCTTGCCGGCCCCGGCCCGGCCTTCAGCCT  557
          |||  ||| | | || || |  |||    | ||| |||   |  | | ||||||||||||||
Db   672  CTACCTGGAGTTCCGGAAGGACCAG--CTAGCCATCCTCCTGCGCCGCTGCTGGAG-AT  728

Qy   558  CTCTCCTGCTCTTTTTTCGCCCTTGGCCATCACAGCCTTCGTCGTACGTGGGCTGCCT  617
          ||||||||||||| ||| || |||| || |||||||||| | ||| | |||||||||
Db   729  GGCGTGGTCCCTCTTCTGTGGTCCCGTGATCATCACCAGCTACTGCTAC--AGCCGCCT  785

Qy   618  CCGGGCACTGCCTGCCTCCGGCAGAGGGGCAGCCACCGCCGCGAGAGGAGGGTGGCGGGGCTGGC  677
          | ||| ||||||   || ||| | |||||| || ||||||||||||||||||||
Db   786  GGTGTGGATCCTCGGCAGAGGGGCAGCCACCGCCGCGAGAGGAGGGTGGCGGGGCTGTGT  845

Qy   678  CGGCGGGGGCCCTCCTCACGCTGCTCTGCGTAGGACCCTACAACGTGGC  737
          |||||||| |||| || | |||| |||| | |||||||||| || |||
Db   678  CGGGGGGCGGCCCTCACGCTGCTCTCTGCGTAGGACCCTACAACGTGGC  737

Qy   738  CAGCTTCCTGTAC  750
          ||||  |||||||
Db   846  GGCGGCCACGCTGCTCAACTTCCTTGTCTGCTTTGGGCCCTACAACGTGTCCCATGTCGT  905

Qy   738  CAGCTTCCTGTAC  750
          |||  ||||||
Db   906  GGGCTATATCTGC  918
```

RESULT 13
US-10-251-385-273
; Sequence 273, Application US/10251385
; Publication No. US20030105292A1
; GENERAL INFORMATION:
;   APPLICANT: Behan, Dominic P.
;   APPLICANT: Chalmers, Derek T.
;   APPLICANT: Liaw, Chen W.
;   TITLE OF INVENTION: No. US2003010105292A1-Endogenous, Constitutively Activated Human G
;   TITLE OF INVENTION: Protein-Coupled
;   TITLE OF INVENTION: Receptors
;   FILE REFERENCE: AREN-0040
;   CURRENT APPLICATION NUMBER: US/10/251,385
;   CURRENT FILING DATE: 2002-09-20
;   PRIOR APPLICATION NUMBER: US/09/170,496
;   PRIOR FILING DATE: 1998-10-13
;   NUMBER OF SEQ ID NOS: 294
;   SOFTWARE: PatentIn version 3.1
; SEQ ID NO 273
;   LENGTH: 1041
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-251-385-273

Query Match   13.1%; Score 118.6; DB 15; Length 1041;
Best Local Similarity   51.8%; Pred. No. 1.5e-19;
Matches 380; Conservative   0; Mismatches 329; Indels   24; Gaps   4;

Qy   18 GCTCTCCTTCGGCCTCTATGTGGCCGCCCTTTGCGCTGGCTTCCCGCTCAACGTCCTGGC 77
          |||||| || |||||| ||||||| ||| |||| | |||| |||||
Db   45 GTTCGTCTTCTCGGTGTACCTTCTCACTTTCCTGTGGGGCTCCCCCTCAACCTGCTGGC 104

Qy   78 CATCCGAGGCGGACGGGCCCACGCGCCCGGCTCCACCCTAGCCGTCTAGCCCCT 137
          |||| |||| | || | ||| |||||||||||| ||||||||||| |||||
Db   105 CCTGGTGGTCTTCGTGGGCAAGCTGCAGCGGCCGGCTCCACCCTAGCTCCTGCT 164

Qy   138 GAACCTGGGCTGCTGCTGCGACCTGTGACAGTCTCTGCCCTGAAGGCGGTGGAGGC 197
          |||||||||||| || || ||||||| ||| || | || || ||||| |||||||||
Db   165 CAACCTGACCGGCCTCGGACCTGTGCGTCCTGTCCTTTCCGCATGGTGGAGGC 224

Qy   198 GCTAGCCTCCGGGGCCTGGGCCCTGGGGGGCTTCGCGTCTTCGGTGTGCCCA 257
          ||||| ||||| |||||| |||| ||| ||| | || ||| ||| |||||||
Db   225 AGCCAATGGCATGCACTGGCCCTGCCCACTCTGCCCACTCTCTGGATTCATCTT 284

Qy   258 CTTCTTCCCACTCTATGCCGCCGCCCTGAGTGAGGACCGCTACCT 317
          ||| | || ||| | ||||| |||||||||| |||||||||||
Db   285 CTTCACCACCATCTATCTCACCGCCCTGCAGCTGTGAGCATTGAACGCTTCCT 344

Qy   318 GGGAGCAGCCTTCCCCCTTGGGCTACCAAGACCCGGCTACCTCCCTGGGGGT 377
          |||||| ||||||| |||| |||||||||| || ||||||||
Db   345 GAGTGTGGCGCTGCTGCTCTGGCTCACCTGGTGCTGACAAGACCCGGCAGGTCTGGT 404
```

Qy 378 GTGCGGCGGCCATCTGGGCCCTCGTCCTGTCTCCTGGGTCTTGGGTCTTTGGGTTGGGAGGC 437
          ||||  ||||| |||||||||  |||| || || |||| |  ||||||||||||||||||
Db 405 GAGTGTGGGCCTGCTGGCTGTGTGGCCTGCTCCACTGCAGCGGTCTCAGTGCTGCTCATAGAATT 464

Qy 438 TCCAGGAGGCTGGCTGGACCACAGCAACACACCTCCCTGGGCATCAACACACCGGTCAACGG 497
          |||||||||                ||||||||||||||||||||||||||||||||||
Db 465 CTCAGGGGG-----------------ACATCTCCCACAGCCAGGGCACCAATGGGACCTG 506

Qy 498 CTCTCCGGTCTGCCTTGGACCCGGACCCCGGCCCTGCCCGCCTTCAGCCT 557
          |||||||||   ||||   || |||||||||| |||||||||||
Db 507 CTACCTGGAGTTCCCGGAAGGACCAG--CTAGCCATCCTCCTGCCCGTGCGCTGGAG-AT 563

Qy 558 CTCTCTCCTGCTCTTTTTCTGCCCTTGGCCATCACAGCCTTCTGCTACGTGGGCTGCCT 617
          ||||      ||||||||   |  ||||||||  |||||||||  ||||
Db 564 GGCTTGTGGTCTCTCTTTGTGTCCCCGCTGATCATCACCAGTACTGCTAC---AGCCGCCT 620

Qy 618 CCGGGCACTGGCCCCTGCCGCTCGACGCACAGGCGGAAGCTGCGGGGCCCGCTGGGTGGC 677
          |||||||||||||   |||| || || ||||||| ||| ||  |||||| |||||
Db 670 TGCCCCTGGACGCACAGGCCCTCCCACTGGCAACCGGCCTTCACCTGGGCTGGTGTGG 929

Qy 548 GCTTCAGCCTCTTCTCCTGCTCTTTTTCTGCCCCTTGGCCATCACAGCCTTCTGCTACG 607
          ||||||||| |||||| || ||||  |||||||   |||| ||||||||||| |||||
Db 930 GCTGTTTCCTGCCCTGCCATGCTGCTGTCTACGGGCACCCCTGCACAGCTGG 989

Qy 608 TGGGCTGCCTCCGGACACTGGCCGCGGGGGCGGCCCGACGCACAGGCGGAAGCGCGGGGCCG 667
          |||||  |||| ||||  || || ||||||||||||||||||  |||||||||  |||||
Db 990 CGGCCAGCGGCCGGCGCTACGGCCACGGCGTGAGGCTGAGGCTGAGGGTGCTGACCCGCTGCTCCG 1049

Qy 668 CCTGGGTGGCCGGCCGGGGGGCCCCTTCTTGCCCCAGCAACCTGCTGCTGCT 704
          |||||||  ||||| ||||||||||||  ||||  ||||||||||
Db 1050 CCGTGGCCCTTCTTCTTCGTGCCCCAGCAGCAACCTGCTGCTGCT 1086

RESULT 14
US-10-225-567A-515
; Sequence 515, Application US/10225567A
; Publication No. US20030113798A1
; GENERAL INFORMATION:
;  APPLICANT: Lifespan Biosciences
;  APPLICANT: Brown, Joseph P.
;  APPLICANT: Burmer, Glenna C.
;  APPLICANT: Roush, Christine L.
;  TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 1920-4-4
;  CURRENT APPLICATION NUMBER: US/10/225,567A
;  CURRENT FILING DATE: 2001-12-19
;  PRIOR APPLICATION NUMBER: 60/257,144
;  PRIOR FILING DATE: 2000-12-19
;  SOFTWARE: PatentIn version 3.1
;  NUMBER OF SEQ ID NOS: 2292
; SEQ ID NO 515
;  LENGTH: 4895
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-10-225-567A-515

  Query Match        9.4%;  Score 84.6;  DB 15;  Length 4895;
  Best Local Similarity  47.2%;  Pred. No. 2.4e-11;
  Matches 329;  Conservative 0;  Mismatches 359;  Indels 9;  Gaps 2;

Qy  8 TGCCCCGCCAGCTCCTCCTTCGGCCTCTATGTGGCCGCCTTGCGCTGGCTTCCCGCTCA 67
          ||||||| ||||| |||  ||  ||||||||||||||||||||||||||
Db 399 TGCCCACCAGGCTGGTGCCCGCCCGCTCTATGGGCCTGGCTGGTGTGGGCGTGCTGCGGCCA 458

Qy 68 ACGTCCTGGCCATCCGAGGGCGACGCGCCCACGCGCCACGCCTCGGTCACCCTAGCCTG 127
          || |||| |||  ||  ||||| |  ||||||||  ||||||| || ||||||
Db 459 ATGGGCTGGGCGTGTGGGGTGCTGGCGGCACCCACCAGGCACCTCGGCGTGCCCCATCGTGC 518

Qy 128 TCTACGCGCCCTGAACCTGGGCTGCTCCGACCTGCTGCTGCTGACAGTCTCTGCCCTGAAGG 187
          |  |||  ||||||||||||||||||||||||||||||  || ||||||  |||||||
Db 519 T------GATGAACCTGCGGCACTGCTGACCTCCTGCTGCCCTGCGCTGCCCCCCGCGGA 572

Qy 188 CGGTGGAGGGCTAGCCTCCGGGGCCTCCTGGCCTCTGCCCTTCGGCCCCTGCGGCCA 247
          ||||| || || ||  |||||||||||||||||||||||||||  ||| |||||
Db 573 TCGCCTACCACCTGCGTGGCCAGCGCTGGCGGCTGGCCCTTCGGGGAGGCCGCCGCCGCCA 632

Qy 248 CGGTGGCCCACTTCTTCCCACTCTATGCCGGCGGGGCGGCCGCTTCCTGCGCCGCCCTGAGTGCAG 307

```
        LENGTH: 4895 base pairs
        TYPE: nucleic acid
        STRANDEDNESS: single
        TOPOLOGY: linear
    FEATURE:
        NAME/KEY: Coding Sequence
        LOCATION: 176...1330
        OTHER INFORMATION:
    SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-187-049-1

Query Match            9.4%;    Score 84.6;  DB 15;   Length 4895;
Best Local Similarity 47.2%;    Pred. No. 2.4e-11;
Matches 329;  Conservative   0;  Mismatches 359;   Indels    9;   Gaps    2;

Qy       8 TGCCCCCGAGCTCTCCTTCGGCCTCTATGTGGCCGCCTTTGCGCTGGGCTTCCCGCTCA  67
             |||| |||        ||  ||||||| ||  || |  ||| || ||||  ||| |||
Db     399 TGCCCACCAGGCTGGTGCCCGCCCTCTATGGCGTCCTGGTCGGTGGGGGCTGCCGGCCA  458

Qy      68 ACGTCCTGGCCATCGAGGCGCGACAGGGCCCACCGCCCGGCTCCGTCTCACCCCTAGCCTGG  127
            ||  ||||        |||||    ||||| |  || |  |||||     |  || |||||
Db     459 ATGGGCTGGCGCTGTGGGTGTCTGGCCACGGAGCACCTCGGCTGCCCTCCACCATGCTGC  518

Qy     128 TCTACGCCCTGAACCTGGCGCTGTTCCGACCTGCTGCTGACAGTCTCTGCCCCTGAAGG  187
            |  |||         || ||||||| ||||| ||||| ||  |  ||||||  |||
Db     519 T-----GATGAACCTGCGACTGCTGACCTCTGCTGCCCGTGCCCTGCCCCGCGGA  572

Qy     188 CGGTGGAGGCGCTAGCCTCCGGGCCTGCCTCTGCCCGGCCTCCGTGCCCGTCTTCG  247
            |     |||            ||||||| ||||||       ||||   ||| |||
Db     573 TCGGCTACCACCTGCGTGGCCAGCGCTGGCCTGCCCTTCGGGAGGCCGCCTGCCCTGGCCA  632

Qy     248 CGGTGGCCCACTTCTTCCCACTCTATGCCGGCGGGGGCTTCCTTGGCCGCCGCCCTGAGTGCAG  307
            |      ||        |    ||  ||||||||  ||    ||||  |  ||||||||
Db     633 CGGCCGCACTCTATGGTCACATGTATGGCTCAGTGCTGCTGCCCGCGTCAGCCTGG  692

Qy     308 GCCGCTACCTGGGAGGAGCAGCCTTCCCCTTGGGCTACCAAGCCTTCCGGAGGCCGTGCTATT  367
            |||||||         || || ||   |      |||||| ||||||  ||||  |  |||
Db     693 ATCGCTACCTGGCCCTGGTGCACCCGGTGCACCCGCCCGCTGCGTGGCCGCGCCTGG  752

Qy     368 CCTGGGGGGTGTGCGCGGCCATCTGGGCCCTCGTCCTGTGTCACCTGGGTCTGTCTTTG  427
            |||            ||||||||      |||   |  || |  |||||| ||| |||
Db     753 CCCTTGGACTTCTGCATGGCTGCTTGG---CTCATGGCGGCGGCCCTGGCACTGCCCCTGA  809

Qy     428 GGTTGGAGGCTCCAGGAGGAGCTGGCTGGACCACCAGCAAACACCTCCCTGGGCATCAAACACAC  487
            ||||||       |      |||| ||      |  ||||  ||  |  ||||| |
Db     810 CACTGCAGCGCAGCAGACCTTCGGCGCGGCCTCCGATCGGCTCTGCTRCTGCCATGACACGCGC  869

Qy     488 CGGTCAACGGCTCTCCCGGTCTGCCCTGGACGTCTTGCCTCGGGAGCCTTCGGCCCGGCCCC  547
            ||    |    |||         |||||| |||||  || |||||||  |  |||||||
Db     870 TGCCCCTGACGACGACAGAGCCTCCACTGGCAACGGGCCTTCAACCTGCCCTTCACCTGCCTTGGCCGCTGTTGG  929

Qy     548 GCTTCAGCCTCCTCTCTCTCCTGCTCTTTTTCTGCCCTTGGCATCACAGCCTTCTGCTACG  607
            |||                | |||||||  |||||| ||||||           ||||
Db     930 GCTGTTTCCTGCCCCCGCTGCTGGCCATGCTGCCATGCTGTCGTCGTGCTACGGGGCCACCCTGCACACGCTGG  989

Qy     608 TGGGCTGCCTCCGGGCACTGGCCCGCTCCGGGCCTGACCGCACAGGCGGAAGCTGCGGGCCG  667
            |       ||       ||||  |||||||||    |||    |  |||| |    |||||
Db     990 CGGCCAGCGACGCCGGGCTACGGCCCACGCGCTGACGGCTGACGCGCAGTGGTGCTGGCCTCCG  1049

Qy     668 CCTGGGTGGCCGGGCCGGGCGGGGCCCTCCTCACGCTGCT  704
            ||   |||             ||  ||||| |||||||||
Db    1050 CCGTGGCCGCCTTCTTCGTGCCCAGCAACCTGCTGCTGCT  1086
```

Search completed: April 30, 2004, 18:51:27
Job time : 468 secs

```
                    GenCore version 5.1.6
                 Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:      April 29, 2004, 12:03:33 ; Search time 59 Seconds
                                         (without alignments)
                                  1436.682 Million cell updates/sec

Title:          US-10-202-687-2
Perfect score:  1584
Sequence:       1 MDLPPQLSFGLYVAAPALGP........RGPGLKTVCAARTQGGKSQK 300

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters:      1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    A_Geneseq_29Jan04:*
           1: geneseqp1980s:*
           2: geneseqp1990s:*
           3: geneseqp2000s:*
           4: geneseqp2001s:*
           5: geneseqp2002s:*
           6: geneseqp2003as:*
           7: geneseqp2003bs:*
           8: geneseqp2004s:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Query Length	DB	ID	Description
1	1584	100.0	300	3	AAY90679	Aay90679 Human G p
2	1584	100.0	300	5	ABG31106	Abg31106 Human G-p
3	1584	100.0	300	6	ABP81994	Abp81994 Human G p
4	1584	100.0	300	7	ADB61404	Adb61404 Human GPR
5	1584	100.0	300	3	ADC22769	Adc22769 Human G p
6	1579	99.7	300	3	AAY90684	Aay90684 Human G p
7	1579	99.7	300	7	ADC22791	Adc22791 Human G p
8	1519	95.9	300	5	ADB61416	Adb61416 Monkey GP
9	1333	84.2	300	5	ABG31107	Abg31107 Mouse G p
10	1333	84.2	300	7	ADB61400	Adb61400 Mouse GPR
11	1320	83.3	300	7	ADB61402	Adb61402 Rat GPR40
12	1295	81.8	300	4	ADB61428	Adb61428 Hamster G
13	366	23.1	346	4	AAB82759	Aab82759 Human G-p
14	347	21.9	346	4	AAY90680	Aay90680 Human G p
15	347	21.9	346	4	AAB82758	Aab82758 Human G-p
16	347	21.9	346	6	ABP81715	Abp81715 Human G p
17	347	21.9	346	2	ADC22773	Adc22773 Human G p
18	347	21.9	401	2	AAW59924	Aaw59924 Human 7-t
19	342	21.6	346	3	AAY90685	Aay90685 Human mut
20	342	21.6	346	3	ADC22793	Adc22793 Human G p
21	340.5	21.5	330	2	AAY05383	Aay05383 Mouse GCR
22	340.5	21.5	330	3	AAY94269	Aay94269 Mouse 7-t
23	334	21.1	330	3	AAY05384	Aay05384 Human GCR
24	332	21.0	330	2	AAW59906	Aaw59906 Human HTA
25	332	21.0	330	3	AAY90681	Aay90681 Human G p.
26	332	21.0	330	6	ABP81990	Abp81990 Human G p
27	332	21.0	330	7	ABR82313	Abr82313 Human G-p
28	332	21.0	330	7	ADC22777	Adc22777 Human G p
29	331	20.9	330	3	AAY90686	Aay90686 Human mut
30	331	20.9	330	7	ADC22795	Adc22795 Human G p
31	322	20.3	330	5	AAW75245	Aaw75245 Fragment
32	322	20.3	330	5	AAE27011	Aae27011 Human gen
33	322	20.3	330	6	AAB27149	Aab27149 Human gen
34	322	20.3	330	6	ABU65022	Abu65022 Human sec
35	298.5	18.8	319	4	AAB82760	Aab82760 Rat G-pro
36	244	15.4	420	6	ABG73510	Abg73510 X. laevis
37	233.5	14.7	388	5	AAG79597	Aag79597 Guinea pi
38	231.5	14.6	385	2	AAY50135	Aay50135 Human pro
39	231.5	14.6	385	2	AAY50139	Aay50139 Mutant hu
40	231.5	14.6	385	2	AAY50138	Aay50138 Mutant hu
41	231.5	14.6	385	2	AAY15082	Aay15082 Human pro
42	231.5	14.6	385	3	AAY45036	Aay45036 Human pro
43	231.5	14.6	385	4	AAB47623	Aab47623 Human PAR
44	231.5	14.6	385	5	AAG79604	Aag79604 Human PAR
45	231.5	14.6	385	6	ABG73512	Abg73512 Human par

ALIGNMENTS

```
RESULT 1
AAY90679
ID   AAY90679 standard; protein; 300 AA.
XX
AC   AAY90679;
XX
DT   21-AUG-2000  (first entry)
XX
DE   Human G protein-coupled receptor GPR40.
XX
KW   G protein-coupled receptor; GPCR; constitutively active;
KW   intracellular loop 3; transmembrane domain 6; drug screening; agonist;
KW   antagonist.
XX
OS   Homo sapiens.
XX
PN   WO20002129-A1.
XX
PD   20-APR-2000.
XX
PF   12-OCT-1999;   99WO-US023938.
XX
PR   13-OCT-1998;   98US-00170496.
XX
PA   (AREN-) ARENA PHARM INC.
XX
PI   Behan DP,  Chalmers DT,  Liaw CW;
XX
DR   WPI; 2000-329165/28.
DR   N-PSDB; AAA30759.
XX
PT   Non-endogenous constitutively activated human G protein-coupled
PT   receptors, useful for identifying agonists for use as pharmaceutical
PT   agents.
XX
PS   Example 1; Page 308-309; 341pp; English.
XX
CC   The invention relates to constitutively active, non-endogenous versions
CC   of endogenous human orphan G protein-coupled receptors (GPCRs, AAY90643-
CC   AAY90677 and AAY90683-Y90687), and to DNA encoding them (AAA30709-A30743
CC   and AAA30775-A30779). The mutant proteins of the invention contain a
CC   mutation in a portion of the protein comprising intracellular loop 3
CC   (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, X,
CC   is substituted for an endogenous residue in IC3 at a position 16 amino
CC   acids N-terminal of an endogenous proline in TM6 to form a sequence X-
CC   (AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg or
CC   Ala, and is preferably Lys. When the endogenous residue at this position
CC   is Lys, this residue is replaced by His, Arg or preferably Ala. The 15
```

```
CC   amino acid stretch between the substituted amino acid and the Pro may be
CC   endogenous, non-endogenous, or a mixture of endogenous and non-endogenous
CC   residues. The constitutively active GPCRs are useful for identifying
CC   antagonists, agonists and partial agonists for use as pharmaceutical
CC   agents. The mutant proteins are also useful in research settings for
CC   elucidating the roles of the receptors in normal and diseased conditions.
CC   Antagonists for a particular GPCR are useful for treating diseases and
CC   disorders associated with that receptor. Because the novel mutant GPCRs
CC   are constitutively active, they can be used directly for screening of
CC   compounds without the need for endogenous ligands. The present sequence
CC   represents a human wild-type GPCR referred to in an exemplification of
CC   the invention
XX
SQ   Sequence 300 AA;

Query Match             100.0%;  Score 1584;  DB 3;  Length 300;
Best Local Similarity   100.0%;  Pred. No. 1.6e-149;
Matches   300;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy          1 MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLP 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLP 60

Qy         61 LKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRP 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRP 120

Qy        121 CYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG 180

Qy        181 PARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY 240

Qy        241 NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK 300


RESULT 2
ABG31106
ID   ABG31106 standard; protein; 300 AA.
XX
AC   ABG31106;
XX
PD   25-JUL-2002.
XX
DT   18-DEC-2001; 2001WO-US048985.
PF   21-OCT-2002 (first entry)
XX
PR   22-DEC-2000; 2000GB-00031527.
XX
DE   Human G protein-coupled receptor GPR40.
XX
KW   Human; GPR40; G protein-coupled; receptor; type 2 diabetes; obesity;
KW   antidiabetic; neuroprotective; anorectic; cerobroprotective; Gq;
KW   G protein; reporter gene; glucose intolerance; insulin intolerance;
KW   neurodegenerative disease; Alzheimer's disease; stroke.
XX
OS   Homo sapiens.
XX
PN   WO200257783-A2.
XX
PD   25-JUL-2002.
XX
PF   18-DEC-2001; 2001WO-US048985.
PR   22-DEC-2000; 2000GB-00031527.
XX
PA   (GLAX ) GLAXO GROUP LTD.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI   Andrews JL, Briscoe CP, Ignar DM, Muir AI, Sauls HR, Tadayyon M;
XX
DR   WPI; 2002-599726/64.
DR   N-PSDB; ABK90236.
XX
```

```
PT   Identifying GPR40 receptor ligand for treating disorders e.g. obesity,
PT   comprises detecting whether the test compound competitively inhibits the
PT   binding of a fatty acid GPR40 ligand to a GPR40 receptor.
XX
PS   Disclosure; Page 48-49; 53pp; English.
XX
CC   The invention relates to screening a test compound to determine whether
CC   the compound is a GPR40 receptor (G protein-coupled receptor) ligand
CC   comprises detecting whether the test compound competitively inhibits the
CC   binding of a fatty acid GPR40 ligand to a GPR40 receptor. Also included
CC   are (1) a method of screening a compound for GPR40 antagonist activity,
CC   comprising: (a) measuring any detectable signal produced by a reporter
CC   gene (comprising a reporter gene under the control of G protein Gq
CC   responsive transcriptional element); (b) detecting a decrease in reporter
CC   gene expression in the presence of both test compound and agonist; or (c)
CC   detecting whether the compound decreases glucose-stimulated insulin
CC   release from mammalian pancreatic beta cells in the presence of a GPR40
CC   agonist, compared to glucose-stimulated insulin release that would occur
CC   due to the presence of the GPR40 agonist; (2) a method of screening a
CC   compound for GPR40 agonist activity, which comprises: (a) detecting any
CC   reporter gene expression; or (b) detecting whether the compound binds to
CC   GPR40 and increases glucose-stimulated insulin release from mammalian
CC   pancreatic beta cells. The method is useful for identifying GPR40
CC   antagonist or agonist compounds for treating disorders e.g. type 2
CC   diabetes, obesity, glucose or insulin intolerance, neurodegenerative
CC   disease (e.g. Alzheimer's disease) or stroke. The present sequence
CC   represents human GPR40
XX
SQ   Sequence 300 AA;

Query Match             100.0%;  Score 1584;  DB 5;  Length 300;
Best Local Similarity   100.0%;  Pred. No. 1.6e-149;
Matches   300;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy          1 MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLP 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLP 60

Qy         61 LKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRP 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRP 120

Qy        121 CYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG 180

Qy        181 PARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY 240

Qy        241 NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK 300


RESULT 3
ABP81994
ID   ABP81994 standard; protein; 300 AA.
XX
AC   ABP81994;
XX
DT   04-MAR-2003 (first entry)
XX
DE   Human G protein-coupled receptor GPR40 protein SEQ ID NO:475.
XX
KW   G protein-coupled receptor; GPCR; antigenic peptide; gene therapy;
KW   G protein-coupled receptor modulator; antibody; immune-related disease;
KW   growth-related disease; cell regeneration-related disease; AIDS; cancer;
KW   immunological-related disease; cell proliferative disease; autoimmune disease;
KW   Alzheimer's disease; atherosclerosis; infection; osteoarthritis; allergy;
KW   osteoporosis; cardiomyopathy; inflammation; Crohn's disease; diabetes;
KW   graft versus host disease; Parkinson's disease; multiple sclerosis; pain;
KW   psoriasis; anxiety; depression; schizophrenia; dementia; memory loss;
```

```
KW    mental retardation; epilepsy; asthma; tuberculosis; obesity; nausea;
KW    hypertension; hypotension; renal disorder; rheumatoid arthritis; trauma;
KW    ulcer.
XX
OS    Homo sapiens.
XX
PN    WO200261087-A2.
XX
PD    08-AUG-2002.
XX
PF    19-DEC-2001; 2001WO-US050107.
XX
PR    19-DEC-2000; 2000US-0257144P.
XX
PA    (LIFE-) LIFESPAN BIOSCIENCES INC.
XX
PI    Burmer GC, Roush CL, Brown JP;
XX
DR    WPI; 2003-046718/04.
DR    N-PSDB; ABZ42843.
XX
PT    New isolated antigenic peptides e.g., for G protein-coupled receptors
PT    (GPCR), useful for diagnosing and designing drugs for treating conditions
PT    in which GPCRs are involved, e.g. AIDS, Alzheimer's disease, cancer or
PT    autoimmune diseases.
XX
PS    Disclosure; Fig 1; 523pp; English.
XX
CC    The present invention describes antigenic peptides (I) comprising: (a)
CC    any one of 1601 sequences (see ABP82019 to ABP83619) of 12-24 amino
CC    acids. Also described: (1) an assay for the detection of a particular G
CC    protein-coupled receptor (GPCR) or a candidate polypeptide in a sample;
CC    and (2) an isolated antibody having high specificity and high affinity or
CC    avidity for a particular GPCR. (I) can be used as GPCR modulators and in
CC    gene therapy. The antigenic peptides for GPCRs are useful in detecting an
CC    antibody against a particular GPCR, and in the production of specific
CC    antibodies. The peptides and antibodies are also useful for detecting the
CC    presence or absence of corresponding GPCRs. The antigenic peptides for
CC    GPCRs and antibodies are useful for diagnosing and designing drugs for
CC    treating immune-related diseases, growth-related diseases, cell
CC    regeneration-related disease, immunological-related cell proliferative
CC    diseases, or autoimmune diseases, e.g. AIDS, Alzheimer's disease,
CC    atherosclerosis, bacterial, fungal, protozoan or viral infections,
CC    osteoarthritis, osteoporosis, cancer, cardiomyopathy, chronic and acute
CC    inflammation, allergies, Crohn's disease, diabetes, graft versus host
CC    disease, Parkinson's disease, multiple sclerosis, pain, psoriasis,
CC    anxiety, depression, schizophrenia, dementia, mental retardation, memory
CC    loss, epilepsy, asthma, tuberculosis, obesity, nausea, hypertension,
CC    hypotension, renal disorders, rheumatoid arthritis, trauma, ulcers, or
CC    any other disorder in which GPCRs are involved. The antibodies may be
CC    used in immunoassays and immunodiagnosis. ABZ42523 to ABZ42869 encode
CC    GPCR proteins given in ABP81675 to ABP82018, which are used in the
CC    exemplification of the present invention
XX
SQ    Sequence 300 AA;

Query Match             100.0%;  Score 1584;  DB 6;  Length 300;
Best Local Similarity   100.0%;
Matches  300;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy          1 MDLPPQLSFGLYVAAFALGPLNVLAIRGATAHARLRLTPSLVVALNLGCSDLLLTVSLP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MDLPPQLSFGLYVAAFALGPLNVLAIRGATAHARLRLTPSLVVALNLGCSDLLLTVSLP 60

Qy         61 LKAVEALASGAWPLPASLCPVFAVAHFPPLYAGGGFLAALSAGRYLGAAPPLGYQAFRRP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKAVEALASGAWPLPASLCPVFAVAHFPPLYAGGGFLAALSAGRYLGAAPPLGYQAFRRP 120

Qy        121 CYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG 180

Qy        181 PARPSLSLLLFFLPLAITTAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY 240
```

```
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PARFSLSLLLFFLPLAITTAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY 240

Qy        241 NASNVASFLYPNLGGSWRKLGLITGAWSVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 NASNVASFLYPNLGGSWRKLGLITGAWSVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK 300

RESULT 4
ADB61404
ID    ADB61404 standard; protein; 300 AA.
XX
AC    ADB61404;
XX
DT    04-DEC-2003  (first entry)
XX
DE    Human GPR40 protein.
XX
KW    fatty acid; eicosanoid-binding G-protein coupled receptor; GPR40;
KW    antidiabetic; anabolic; neuroprotective; nephrotropic; ophthalmological;
KW    antilipaemic; dermatological; antiarteriosclerotic; antiarthritic;
KW    osteopathic; thrombolytic; nootropic; anorectic; hypotensive; cytostatic;
KW    diabetes; ketosis; acidosis; diabetic neuropathy; diabetic nephropathy;
KW    diabetic retinopathy; hyperlipaemia; skin disease; arthritis;
KW    bone disease; arteriosclerosis; thrombosis; hypertension; indigestion;
KW    memory loss; obesity; hypoglycaemia; edema; insulin resistance;
KW    insulin allergy; fat metabolism disorder; cancer; human.
XX
OS    Homo sapiens.
XX
PN    WO200306859-A1.
XX
PD    21-AUG-2003.
XX
PP    13-FEB-2003; 2003WO-JP001483.
XX
PR    14-FEB-2002; 2002JP-00037131.
PR    12-JUL-2002; 2002JP-00204163.
PR    12-NOV-2002; 2002JP-00328696.
PR    22-JAN-2003; 2003JP-00014032.
XX
PA    (TAKE ) TAKEDA CHEM IND LTD.
XX
PI    Hinuma S, Hosoya M, Ito Y, Kobayashi M, Tanaka H, Okubo S;
PI    Fujii R, Kizawa H, Kawamata Y, Ogi K, Harada M, Fukusumi S;
XX
DR    WPI; 2003-671661/63.
DR    N-PSDB; ADB61403.
XX
PT    Fatty acid and eicosanoid-binding G-protein coupled receptor protein
PT    GPR40 for control of pancreatic function and treatment of diabetes.
XX
PS    Claim 34; Page 235-236; 257pp; Japanese.
XX
CC    The invention relates to a novel screening method comprising a fatty acid
CC    and eicosanoid-binding G-protein coupled receptor protein, GPR40,
CC    originating in mouse, rat, crab-eating monkey and hamster (sequences
CC    fully defined in the specification), equivalent proteins of similar
CC    activity, and peptides containing partial sequences of the GPR40 protein.
CC    The novel fatty acid and GPR40 protein and their compositions have the
CC    following activities: antidiabetic, anabolic, neuroprotective,
CC    nephrotropic, ophthalmological, antilipaemic, dermatological,
CC    antiarteriosclerotic, antiarthritic, osteopathic, thrombolytic,
CC    nootropic, anorectic, hypotensive, and cytostatic. The novel fatty acid
CC    and eicosanoid-binding G-protein coupled receptor protein, GPR40, and
CC    further compositions can be used in the treatment, prevention and
CC    diagnosis of diabetes, ketosis, acidosis, diabetic neuropathy, diabetic
CC    nephropathy, diabetic retinopathy, hyperlipaemia, skin diseases,
CC    arthritis, bone diseases, arteriosclerosis, thrombosis, hypertension,
CC    indigestion, memory loss, obesity, hypoglycaemia, edema, insulin
CC    resistance, insulin allergy, fat metabolism disorders and cancer. This
CC    sequence represents a human GPR40 protein of the invention.
XX
```

Left column

```
SQ   Sequence 300 AA;

     Query Match           100.0%;    Score 1584;  DB 7;   Length 300;
     Best Local Similarity 100.0%;    Pred. No. 1.6e-149;
     Matches 300;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy     1 MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLP 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLP 60

Qy    61 LKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRP 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 LKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRP 120

Qy   121 CYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG 180

Qy   181 PARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 PARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY 240

Qy   241 NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK 300

RESULT 5
ADC22769
ID   ADC22769 standard; protein; 300 AA.
XX
AC   ADC22769;
XX
DT   18-DEC-2003 (first entry)
XX
DE   Human G protein-coupled receptor (GPCR) polypeptide #36.
XX
KW   Human; G protein-coupled receptor; GPCR; transmembrane-6 region; TM6;
KW   intracellular-3 region; IC3; receptor.
XX
OS   Homo sapiens.
XX
PN   US6555339-B1.
XX
PD   29-APR-2003.
XX
PF   13-OCT-1998;  98US-00170496.
XX
PR   14-APR-1997;  97US-00839449.
PR   14-APR-1998;  98US-00060188.
PR   26-JUN-1998;  98US-0090783P.
PR   07-AUG-1998;  98US-0095677P.
XX
PA   (AREN-) ARENA PHARM INC.
XX
PI   Liaw CW, Behan DP, Chalmers DT;
XX
DR   WPI; 2003-742861/70.
DR   N-PSDB; ADC22768.
XX
PT   Creating a constitutively active version of an endogenous human G protein
PT   coupled receptor (GPCR) comprises substituting a specific amino acid in
PT   the transmembrane-6 region with a different amino acid, and testing for
PT   constitutive activity.
XX
PS   Example 1; SEQ ID NO 250; 221pp; English.
XX
CC   The invention relates to a method for treating a non-endogenous,
CC   constitutively active version of an endogenous human G protein-coupled
CC   receptor (GPCR) that has a transmembrane-6 (TM6) region and an
CC   intracellular-3 (IC3) region, by substituting a specific amino acid in
CC   the TM6 region with a different amino acid, and testing for constitutive
CC   activity. The method is useful for creating a constitutively active
```

Right column

```
CC   version of an endogenous human GPCR that comprises a transmembrane 6
CC   region and an intracellular loop 3 region. The altered human GPCR
CC   polypeptides are useful for screening test compounds for identification
CC   of inverse agonists or partial agonists of GPCR polypeptides, which may
CC   have therapeutic uses. The altered GPCRs may also be used in vivo or in
CC   vitro in biological research. A nucleic acid encoding the altered GPCR
CC   may be used to create a transgenic animal expressing the altered GPCR.
CC   The method allows screening for compounds that modulate the activity of a
CC   human G protein-coupled receptor without the need for provision of a
CC   ligand for the receptor. This is particularly useful in allowing
CC   screening of compounds against orphan receptors for which no ligand is
CC   currently known. This sequence represents a human GPCR polypeptide of the
CC   invention.
XX
SQ   Sequence 300 AA;

     Query Match           100.0%;    Score 1584;  DB 7;   Length 300;
     Best Local Similarity 100.0%;    Pred. No. 1.6e-149;
     Matches 300;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy     1 MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLP 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLP 60

Qy    61 LKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRP 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 LKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRP 120

Qy   121 CYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG 180

Qy   181 PARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 PARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY 240

Qy   241 NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK 300

RESULT 6
AAY90684
ID   AAY90684 standard; protein; 300 AA.
XX
AC   AAY90684;
XX
DT   21-AUG-2000 (first entry)
XX
DE   Human mutant G protein-coupled receptor GPR40 (A223K).
XX
KW   G protein-coupled receptor; GPCR; constitutively active;
KW   intracellular loop 3; transmembrane domain 6; drug screening; agonist;
KW   antagonist; mutant; mutein.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO20002129-A1.
XX
PD   20-APR-2000.
XX
PF   21-AUG-2000 (first entry)
XX
PR   12-OCT-1999;  99WO-US023938.
XX
PR   13-OCT-1998;  98US-00170496.
XX
PA   (AREN-) ARENA PHARM INC.
XX
PI   Behan DP, Chalmers DT, Liaw CW;
XX
DR   WPI; 2000-329165/28.
DR   N-PSDB; AAA30776.
XX
```

PT   Non-endogenous constitutively activated human G protein-coupled
PT   receptors, useful for identifying agonists for use as pharmaceutical
PT   agents.
XX
PS   Example 2; Page 325-326; 341pp; English.
XX
CC   The invention relates to constitutively active, non-endogenous versions
CC   of endogenous human orphan G protein-coupled receptors (GPCRs, AAY90643-
CC   AAY90677 and AAY90683-Y90687), and to DNA encoding them (AAA30709-A30743
CC   and AAA30775-A30779). The mutant proteins of the invention contain a
CC   mutation in a portion of the protein comprising intracellular loop 3
CC   (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, X,
CC   is substituted for an endogenous residue in IC3 at a position 16 amino
CC   acids N-terminal of an endogenous proline in TM6 to form a sequence X-
CC   (AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg or
CC   Ala, and is preferably Lys. When the endogenous residue at this position
CC   is Lys, this residue is replaced by His, Arg or preferably Ala. The 15
CC   amino acid stretch between the substituted amino acid and the Pro may be
CC   endogenous, non-endogenous, or a mixture of endogenous and non-endogenous
CC   residues. The constitutively active GPCRs are useful for identifying
CC   antagonists, agonists and partial agonists for use as pharmaceutical
CC   agents. The mutant proteins are also useful in research settings for
CC   elucidating the roles of the receptors in normal and diseased conditions.
CC   Antagonists for a particular GPCR are useful for treating diseases and
CC   disorders associated with that receptor. Because the novel mutant GPCRs
CC   are constitutively active, they can be used directly for screening of
CC   compounds without the need for endogenous ligands. Sequences AAY90643-
CC   AAY90683-Y90687 the mutant human GPCRs of the invention
XX
SQ   Sequence 300 AA;

Query Match   99.7%; Score 1579; DB 3; Length 300;
Best Local Similarity 99.7%; Pred. No. 5.1e-149;
Matches 299; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy   1 MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLP 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLP 60

Qy   61 LKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRP 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRP 120

Qy 121 CYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 CYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG 180

Qy 181 PARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY 240
         |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db 181 PARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAKWVAGGALLTLLLCVGPY 240

Qy 241 NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK 300

RESULT 7
ADC22791
ID   ADC22791 standard; protein; 300 AA.
XX
AC   ADC22791;
XX
DT   18-DEC-2003 (first entry)
XX
DE   Human G protein-coupled receptor (GPCR) polypeptide #76.
XX
KW   Human; G protein-coupled receptor; GPCR; transmembrane-6 region; TM6;
KW   intracellular-3 region; IC3; receptor.
XX
OS   Homo sapiens.
XX
PN   US655339-B1.
XX

PD   29-APR-2003.
XX
PF   13-OCT-1998; 98US-00170496.
XX
PR   14-APR-1997; 97US-00839449.
PR   14-APR-1998; 98US-00060188.
PR   26-JUN-1998; 98US-0090783P.
PR   07-AUG-1998; 98US-0095677P.
PA   (AREN-) ARENA PHARM INC.
XX
PI   Liaw CW, Behan DP, Chalmers DT;
XX
DR   WPI; 2003-742861/70.
DR   N-PSDB; ADC22790.
XX
PT   Creating a constitutively active version of an endogenous human G protein
PT   coupled receptor (GPCR) comprises substituting a specific amino acid in
PT   the transmembrane-6 region with a different amino acid, and testing for
PT   constitutive activity.
XX
PS   Example 2; SEQ ID NO 272; 221pp; English.
XX
CC   The invention relates to a method for treating a non-endogenous,
CC   constitutively active version of an endogenous human G protein-coupled
CC   receptor (GPCR) that has a transmembrane-6 (TM6) region and an
CC   intracellular-3 (IC3) region, by substituting a specific amino acid in
CC   the TM6 region with a different amino acid, and testing for constitutive
CC   activity. The method is useful for creating a constitutively active
CC   version of an endogenous human GPCR that comprises a transmembrane 6
CC   region and an intracellular loop 3 region. The altered human GPCR
CC   polypeptides are useful for screening test compounds for identification
CC   of inverse agonists or partial agonists of GPCR polypeptides, which may
CC   have therapeutic uses. The altered GPCRs may also be used in vivo or in
CC   vitro in biological research. A nucleic acid encoding the altered GPCR
CC   may be used to create a transgenic animal expressing the altered GPCR.
CC   The method allows screening for compounds that modulate the activity of a
CC   human G protein-coupled receptor without the need for provision of a
CC   ligand for the receptor. This is particularly useful in allowing
CC   screening of compounds against orphan receptors for which no ligand is
CC   currently known. This sequence represents a human GPCR polypeptide of the
CC   invention.
XX
SQ   Sequence 300 AA;

Query Match   99.7%; Score 1579; DB 7; Length 300;
Best Local Similarity 99.7%; Pred. No. 5.1e-149;
Matches 299; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy   1 MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLP 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLP 60

Qy   61 LKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRP 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRP 120

Qy 121 CYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 CYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG 180

Qy 181 PARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY 240
         |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db 181 PARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAKWVAGGALLTLLLCVGPY 240

Qy 241 NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK 300

RESULT 8
ADB61416
ID   ADB61416 standard; protein; 300 AA.

```
XX
AC   ADB61416;
XX
DT   04-DEC-2003  (first entry)
XX
DE   Monkey GPR40 protein.
XX
KW   fatty acid; eicosanoid-binding G-protein coupled receptor; GPR40;
KW   antidiabetic; anabolic; neuroprotective; nephrotropic; ophthalmological;
KW   antilipaemic; dermatological; antiarteriosclerotic; antiarthritic;
KW   osteopathic; thrombolytic; nootropic; anorectic; hypotensive; cytostatic;
KW   diabetes; ketosis; acidosis; diabetic neuropathy; diabetic nephropathy;
KW   diabetic retinopathy; hyperlipaemia; skin disease; arthritis;
KW   bone disease; arteriosclerosis; thrombosis; hypertension; indigestion;
KW   memory loss; obesity; hypoglycaemia; edema; insulin resistance;
KW   insulin allergy; fat metabolism disorder; cancer; monkey.
XX
OS   Primates.
XX
PN   WO2003068959-A1.
XX
PD   21-AUG-2003.
XX
PF   13-FEB-2003; 2003WO-JP001483.
XX
PR   14-FEB-2002; 2002JP-00037131.
PR   12-JUL-2002; 2002JP-00204163.
PR   12-NOV-2002; 2002JP-00328696.
PR   22-JAN-2003; 2003JP-00014032.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Hinuma S, Hosoya M, Ito Y, Kobayashi M, Tanaka H, Okubo S;
PI   Fujii R, Kizawa H, Kawamata Y, Ogi K, Harada M, Fukusumi S;
XX
DR   WPI; 2003-671661/63.
DR   N-PSDB; ADB61417.
XX
PT   Fatty acid and eicosanoid-binding G-protein coupled receptor protein
PT   GPR40 for control of pancreatic function and treatment of diabetes.
XX
PS   Claim 1; Page 240-241; 257pp; Japanese.
XX
CC   The invention relates to a novel screening method comprising a fatty acid
CC   and eicosanoid-binding G-protein coupled receptor protein, GPR40,
CC   originating in mouse, rat, crab-eating monkey and hamster (sequences
CC   fully defined in the specification), equivalent proteins of similar
CC   activity, and peptides containing partial sequences of the GPR40 protein.
CC   The novel fatty acid and GPR40 protein and their compositions have the
CC   following activities: antidiabetic, anabolic, neuroprotective,
CC   nephrotropic, ophthalmological, antilipaemic, dermatological,
CC   antiarteriosclerotic, antiarthritic, osteopathic, thrombolytic,
CC   nootropic, anorectic, hypotensive, and cytostatic. The novel fatty acid
CC   and eicosanoid-binding G-protein coupled receptor protein, GPR40, and
CC   further compositions can be used in the treatment, prevention and
CC   diagnosis of diabetes, ketosis, acidosis, diabetic neuropathy, diabetic
CC   nephropathy, diabetic retinopathy, hyperlipaemia, skin diseases,
CC   arthritis, bone diseases, arteriosclerosis, thrombosis, hypertension,
CC   indigestion, memory loss, obesity, hypoglycaemia, edema, insulin
CC   resistance, insulin allergy, fat metabolism disorders and cancer. This
CC   sequence represents a monkey GPR40 protein of the invention.
XX
SQ   Sequence 300 AA;

Query Match             95.9%;  Score 1519;  DB 7;  Length 300;
Best Local Similarity   96.7%;  Pred. No. 4.9e-143;
Matches  290;  Conservative    0;  Mismatches  10;  Indels    0;  Gaps    0;

Qy          1 MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLP 60
              ||||||||| |||||||||||||||||||| ||||| |||||||||||||||||||||||
Db          1 MDLPPQLSFALYVAAFALGFPLNVLAIRGARAHARRRLTPSLVYALNLGCSDLLLTVSLP 60

Qy         61 LKAVEALASGAWPLPASLCPVFGVAHFAPLYAGGGFLAALSAGRYLGAAFFPLGYQAFRRP 120
```

```
Db         61 LKAVEALASGAWPLPASLCPVFGVAHFAPLYAGGGFLAALSAGRYLGAAFFPLGYQAFRRP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy        121 CYSNGVCAAIWALVLCHLGLVFVLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CYSNGVCAAIWALVLCHLGLVFVLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG 180
Qy        181 PARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY 240
              ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db        181 PARFSLSLLLFFLPLAITAFCYVGCLRALAHSGLTHRRKLRAAWVAGGALLTLLLCVGPY 240
Qy        241 NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK 300
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||| ||||
Db        241 NASNVASFLNPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGSTSQK 300

RESULT 9
ABG31107
ID   ABG31107 standard; protein; 300 AA.
XX
AC   ABG31107;
XX
DT   21-OCT-2002  (first entry)
XX
DE   Mouse G protein-coupled receptor GPR40.
XX
KW   Mouse; GPR40; G protein-coupled; receptor; type 2 diabetes; obesity;
KW   antidiabetic; neuroprotective; anorectic; cerobroprotective; Gq;
KW   G protein; reporter gene; glucose intolerance; insulin intolerance;
KW   neurodegenerative disease; Alzheimer's disease; stroke.
XX
OS   Mus sp.
XX
PN   WO200257783-A2.
XX
PD   25-JUL-2002.
XX
PF   18-DEC-2001; 2001WO-US048985.
PR   22-DEC-2000; 2000GB-00031527.
XX
PA   (GLAX ) GLAXO GROUP LTD.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI   Andrews JL,  Briscoe CP,  Ignar DM,  Muir AI,  Sauls HR,  Tadayyon M;
XX
DR   WPI; 2002-599726/64.
DR   N-PSDB; ABK90237.
XX
PT   Identifying GPR40 receptor ligand for treating disorders e.g. obesity,
PT   comprises detecting whether the test compound competitively inhibits the
PT   binding of a fatty acid GPR40 ligand to a GPR40 receptor.
XX
PS   Disclosure; Page 51; 53pp; English.
XX
CC   The invention relates to screening a test compound to determine whether
CC   gene expression in the presence of both test compound and agonist, or (c)
CC   the compound is a GPR40 receptor (G protein-coupled receptor) ligand
CC   comprises detecting whether the test compound competitively inhibits the
CC   binding of a fatty acid GPR40 ligand to a GPR40 receptor. Also included
CC   are (1) a method of screening a compound for GPR40 antagonist activity,
CC   comprising: (a) measuring any detectable signal produced by a reporter
CC   gene (comprising a reporter gene under the control of G protein Gq
CC   responsive transcriptional element); (b) detecting a decrease in reporter
CC   detecting whether the compound decreases glucose-stimulated insulin
CC   release from mammalian pancreatic beta cells in the presence of a GPR40
CC   agonist, compared to glucose-stimulated insulin release that would occur
CC   due to the presence of the GPR40 agonist; (2) a method of screening a
CC   compound for GPR40 agonist activity, which comprises: (a) detecting any
CC   reporter gene expression; or (b) detecting whether the compound binds to
CC   GPR40 and increases glucose-stimulated insulin release from mammalian
CC   pancreatic beta cells. The method is useful for identifying GPR40
CC   antagonist or agonist compounds for treating disorders e.g. type 2
```

Left column

```
CC  diabetes, obesity, glucose or insulin intolerance, neurodegenerative
CC  disease (e.g. Alzheimer's disease) or stroke. The present sequence
CC  represents mouse GPR40
XX
SQ  Sequence 300 AA;

Query Match             84.2%;  Score 1333;  DB 5;  Length 300;
Best Local Similarity   83.0%;  Pred. No. 1.8e-124;
Matches  249;  Conservative  17;  Mismatches  34;  Indels  0;  Gaps  0;

Qy      1 MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLP  60
          |||||||| ||| :|||||||||  |||||  :|| :|||||||| : |||||||| ::|
Db      1 MDLPPQLSFALYVSAFALGFPLNLLAIRGAVSHAKLRLTPSLVYTLHLGCSDLLLAITLP  60

Qy     61 LKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRP 120
          |||||||||||||   |  ||||| ||| |||||||||||||||||||||||  ||||  || 
Db     61 LKAVEALASGAWPLPLPFCPVFALAHFAPLYAGGGFLAALSAGRYLGAAFPPGYQAIRRP 120

Qy    121 CYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG 180
          ||||||| |||||||||||| | | :: |||| |:|||| ||||||||||||||||| |
Db    121 RYSWGVCVAIWALVLCHLGLALGLETSGSWLDNSTSSLGINIPVNGSPVCLEAWDPDSAR 180

Qy    181 PARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY 240
          |||  |||||||||| ||||||||||||| ||||: ||||:|||||||||||||||| ||| 
Db    181 PARLSFSILLFFLPLVITAFCYVGCLRALVRSGLSHKRKLRAAWVAGGALLTLLLCLGPY 240

Qy    241 NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTGGGKSQK 300
          |||||||: |   |||||||||||||||||||||||||||| :|  :|  :|  :
Db    241 NASNVASFINPDLGGSWRKLGLITGAWSVVLNPLVTGYLGTGPGRGTICVTRTQRGTIQK 300
```

```
RESULT 10
ADB61400
ID  ADB61400 standard; protein; 300 AA.
XX
AC  ADB61400;
XX
DT  04-DEC-2003  (first entry)
XX
DE  Mouse GPR40 protein.
XX
KW  fatty acid; eicosanoid-binding G-protein coupled receptor; GPR40;
KW  antidiabetic; anabolic; neuroprotective; nephrotropic; ophthalmological;
KW  antilipaemic; dermatological; antiarteriosclerotic; antiarthritic;
KW  osteopathic; thrombolytic; nootropic; anorectic; hypotensive; cytostatic;
KW  diabetes; ketosis; acidosis; diabetic neuropathy; diabetic nephropathy;
KW  diabetic retinopathy; hyperlipaemia; skin disease; arthritis;
KW  bone disease; arteriosclerosis; thrombosis; hypertension; indigestion;
KW  memory loss; obesity; hypoglycaemia; edema; insulin resistance;
KW  insulin allergy; fat metabolism disorder; cancer; mouse; murine.
XX
OS  Mus sp.
XX
PN  WO2003068959-A1.
XX
PD  21-AUG-2003.
XX
PF  13-FEB-2003; 2003WO-JP001483.
XX
PR  14-FEB-2002; 2002JP-00037131.
PR  12-JUL-2002; 2002JP-00204163.
PR  12-NOV-2002; 2002JP-00328696.
PR  22-JAN-2003; 2003JP-00014032.
XX
PA  (TAKE ) TAKEDA CHEM IND LTD.
XX
PI  Hinuma S,  Hosoya M,  Ito Y,  Kobayashi M,  Tanaka H,  Okubo S;
PI  Fujii R,  Kizawa H,  Kawamata Y,  Ogi K,  Harada M,  Fukusumi S;
XX
DR  WPI; 2003-671661/63.
DR  N-PSDB; ADB61401.
XX
PT  Fatty acid and eicosanoid-binding G-protein coupled receptor protein
```

Right column

```
PT  GPR40 for control of pancreatic function and treatment of diabetes.
XX
PS  Claim 1; Page 230-231; 257pp; Japanese.
XX
CC  The invention relates to a novel screening method comprising a fatty acid
CC  and eicosanoid-binding G-protein coupled receptor protein, GPR40,
CC  originating in mouse, rat, crab-eating monkey and hamster (sequences
CC  fully defined in the specification), equivalent proteins of similar
CC  activity, and peptides containing partial sequences of the GPR40 protein.
CC  The novel fatty acid and GPR40 protein and their compositions have the
CC  following activities: antidiabetic, anabolic, neuroprotective,
CC  nephrotropic, ophthalmological, antilipaemic, dermatological,
CC  antiarteriosclerotic, antiarthritic, osteopathic, thrombolytic,
CC  nootropic, anorectic, hypotensive, and cytostatic. The novel fatty acid
CC  and eicosanoid-binding G-protein coupled receptor protein, GPR40, and
CC  further compositions can be used in the treatment, prevention and
CC  diagnosis of diabetes, ketosis, acidosis, diabetic neuropathy, diabetic
CC  nephropathy, diabetic retinopathy, hyperlipaemia, skin diseases,
CC  arthritis, bone diseases, arteriosclerosis, thrombosis, hypertension,
CC  indigestion, memory loss, obesity, hypoglycaemia, edema, insulin
CC  resistance, insulin allergy, fat metabolism disorders and cancer. This
CC  sequence represents a mouse GPR40 protein of the invention.
XX
SQ  Sequence 300 AA;

Query Match             84.2%;  Score 1333;  DB 7;  Length 300;
Best Local Similarity   83.0%;  Pred. No. 1.8e-124;
Matches  249;  Conservative  17;  Mismatches  34;  Indels  0;  Gaps  0;

Qy      1 MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLP  60
          |||||||| ||| :|||||||||  |||||  :|| :|||||||| : |||||||| ::|
Db      1 MDLPPQLSFALYVSAFALGFPLNLLAIRGAVSHAKLRLTPSLVYTLHLGCSDLLLAITLP  60

Qy     61 LKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRP 120
          |||||||||||||   |  ||||| ||| |||||||||||||||||||||||  ||||  || 
Db     61 LKAVEALASGAWPLPLPFCPVFALAHFAPLYAGGGFLAALSAGRYLGAAFPPGYQAIRRP 120

Qy    121 CYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG 180
          ||||||| |||||||||||| | | :: |||| |:|||| ||||||||||||||||| |
Db    121 RYSWGVCVAIWALVLCHLGLALGLETSGSWLDNSTSSLGINIPVNGSPVCLEAWDPDSAR 180

Qy    181 PARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY 240
          |||  |||||||||| ||||||||||||| ||||: ||||:|||||||||||||||| ||| 
Db    181 PARLSFSILLFFLPLVITAFCYVGCLRALVRSGLSHKRKLRAAWVAGGALLTLLLCLGPY 240

Qy    241 NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTGGGKSQK 300
          |||||||: |   |||||||||||||||||||||||||||| :|  :|  :|  :
Db    241 NASNVASFINPDLGGSWRKLGLITGAWSVVLNPLVTGYLGTGPGRGTICVTRTQRGTIQK 300
```

```
RESULT 11
ADB61402
ID  ADB61402 standard; protein; 300 AA.
XX
AC  ADB61402;
XX
DT  04-DEC-2003  (first entry)
XX
DE  Rat GPR40 protein.
XX
KW  fatty acid; eicosanoid-binding G-protein coupled receptor; GPR40;
KW  antidiabetic; anabolic; neuroprotective; nephrotropic; ophthalmological;
KW  antilipaemic; dermatological; antiarteriosclerotic; antiarthritic;
KW  osteopathic; thrombolytic; nootropic; anorectic; hypotensive; cytostatic;
KW  diabetes; ketosis; acidosis; diabetic neuropathy; diabetic nephropathy;
KW  diabetic retinopathy; hyperlipaemia; skin disease; arthritis;
KW  bone disease; arteriosclerosis; thrombosis; hypertension; indigestion;
KW  memory loss; obesity; hypoglycaemia; edema; insulin resistance;
KW  insulin allergy; fat metabolism disorder; cancer; rat.
XX
OS  Rattus sp.
XX
PN  WO2003068959-A1.
```

Mon May 3 09:27:00 2004

us-10-202-687-2.rag

Page 8

XX
PD   21-AUG-2003.
XX
PF   13-FEB-2003; 2003WO-JP001483.
XX
PR   14-FEB-2002; 2002JP-00037131.
PR   12-JUL-2002; 2002JP-00204163.
PR   12-NOV-2002; 2002JP-00328696.
PR   22-JAN-2003; 2003JP-00014032.
XX
PA   (TAKE) TAKEDA CHEM IND LTD.
XX
PI   Hinuma S, Hosoya M, Ito Y, Kobayashi M, Tanaka H, Okubo S;
PI   Fujii R, Kizawa H, Kawamata Y, Ogi K, Harada M, Fukusumi S;
XX
DR   WPI; 2003-671661/63.
DR   N-PSDB; ADB61403.
XX
PT   Fatty acid and eicosanoid-binding G-protein coupled receptor protein
PT   GPR40 for control of pancreatic function and treatment of diabetes.
XX
PS   Claim 1; Page 232-234; 257pp; Japanese.
XX
CC   The invention relates to a novel screening method comprising a fatty acid
CC   and eicosanoid-binding G-protein coupled receptor protein, GPR40,
CC   originating in mouse, rat, crab-eating monkey and hamster (sequences
CC   fully defined in the specification), equivalent proteins of similar
CC   activity, and peptides containing partial sequences of the GPR40 protein.
CC   The novel fatty acid and GPR40 protein and their compositions have the
CC   following activities: antidiabetic, anabolic, neuroprotective,
CC   nephrotropic, ophthalmological, antilipaemic, antidiabetic,
CC   antiarteriosclerotic, antiarthritic, osteopathic, thrombolytic,
CC   nootropic, anorectic, hypotensive, and cytostatic. The novel fatty acid
CC   and eicosanoid-binding G-protein coupled receptor protein, GPR40, and
CC   further compositions can be used in the treatment, prevention and
CC   diagnosis of diabetes, ketosis, acidosis, diabetic neuropathy, diabetic
CC   nephropathy, diabetic retinopathy, hyperlipaemia, skin diseases,
CC   arthritis, bone diseases, arteriosclerosis, thrombosis, hypertension,
CC   indigestion, memory loss, obesity, hypoglycaemia, edema, insulin
CC   resistance, insulin allergy, fat metabolism disorders and cancer. This
CC   sequence represents a rat GPR40 protein of the invention.
XX
SQ   Sequence 300 AA;

Query Match   83.3%; Score 1320; DB 7; Length 300;
Best Local Similarity   81.7%; Pred. No. 3.6e-123;
Matches 245; Conservative 19; Mismatches 36; Indels 0; Gaps 0;

Qy   1 MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLP 60
            ||||||||||||:||||:|||||||||||::|||:||||||||||||||||||||::||
Db   1 MDLPPQLSFALYVSAFALGFPLNLLAIRGAVSHAKLRLTPSLVVTLHLACSDLLLAITLP 60

Qy   61 LKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRP 120
            |||||||||:||||| |||||||||||||||||||||||||||||||||||||||||
Db   61 LKAVEALASGVWPLPLPFCPVFALAHFAPLYAGGGFLAALSAGRYLGAAFPFGYQAIRRP 120

Qy   121 CYSWNGVCAAIWALVLCHLGLVFGGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG 180
            ||||::|:||||||||||||:||:|::|||||||| | :|:||:|||||||
Db   121 CYSWGVCVAIWALVLCHLGLALGLEAPRGWVDNTTSSLGINIPVNGSPVCLEAWDPDSAR 180

Qy   181 PARFSLSLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY 240
            ||||:|:|||||||:|||||||||||||||:||||:||:|||||||||||||||||||
Db   181 PARLSFSTILLFFLPLVITAFCYVGCLRALVHSGLSHKRKLRAAWVAGGALLTLLLCLGPY 240

Qy   241 NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLLVTGYLGRGPGLKTVCAARTQGGKSQK 300
            ||||||||:|||||||||||||||||||||||:|:|||||||:|:|| :||:|||||
Db   241 NASNVASFINPDLEGSWRKLGLITGAWSVVLNPLVTGYLGTGPGQGTICVTRTPRGTIQK 300

RESULT 12
ADB61428
ID   ADB61428 standard; protein; 300 AA.
XX

AC   ADB61428;
XX
DT   04-DEC-2003 (first entry)
XX
DE   Hamster GPR40 protein.
XX
KW   fatty acid; eicosanoid-binding G-protein coupled receptor; GPR40;
KW   antidiabetic; anabolic; neuroprotective; nephrotropic; ophthalmological;
KW   antilipaemic; dermatological; antiarteriosclerotic; antiarthritic;
KW   osteopathic; thrombolytic; nootropic; anorectic; hypotensive; cytostatic;
KW   diabetes; ketosis; acidosis; diabetic neuropathy; diabetic nephropathy;
KW   diabetic retinopathy; hyperlipaemia; skin disease; arthritis;
KW   bone disease; arteriosclerosis; thrombosis; hypertension; indigestion;
KW   memory loss; obesity; hypoglycaemia; edema; insulin resistance;
KW   insulin allergy; fat metabolism disorder; cancer; hamster.
XX
OS   Cricetinae gen. sp.
XX
PN   WO200306959-A1.
XX
PD   21-AUG-2003.
XX
PF   13-FEB-2003; 2003WO-JP001483.
XX
PR   14-FEB-2002; 2002JP-00037131.
PR   12-JUL-2002; 2002JP-00204163.
PR   12-NOV-2002; 2002JP-00328696.
PR   22-JAN-2003; 2003JP-00014032.
XX
PA   (TAKE) TAKEDA CHEM IND LTD.
XX
PI   Hinuma S, Hosoya M, Ito Y, Kobayashi M, Tanaka H, Okubo S;
PI   Fujii R, Kizawa H, Kawamata Y, Ogi K, Harada M, Fukusumi S;
XX
DR   WPI; 2003-671661/63.
DR   N-PSDB; ADB61429.
XX
PT   Fatty acid and eicosanoid-binding G-protein coupled receptor protein
PT   GPR40 for control of pancreatic function and treatment of diabetes.
XX
PS   Claim 1; Page 245-246; 257pp; Japanese.
XX
CC   The invention relates to a novel screening method comprising a fatty acid
CC   and eicosanoid-binding G-protein coupled receptor protein, GPR40,
CC   originating in mouse, rat, crab-eating monkey and hamster (sequences
CC   fully defined in the specification), equivalent proteins of similar
CC   activity, and peptides containing partial sequences of the GPR40 protein.
CC   The novel fatty acid and GPR40 protein and their compositions have the
CC   following activities: antidiabetic, anabolic, neuroprotective,
CC   nephrotropic, ophthalmological, antilipaemic, dermatological,
CC   antiarteriosclerotic, antiarthritic, osteopathic, thrombolytic,
CC   nootropic, anorectic, hypotensive, and cytostatic. The novel fatty acid
CC   and eicosanoid-binding G-protein coupled receptor protein, GPR40, and
CC   further compositions can be used in the treatment, prevention and
CC   diagnosis of diabetes, ketosis, acidosis, diabetic neuropathy, diabetic
CC   nephropathy, diabetic retinopathy, hyperlipaemia, skin diseases,
CC   arthritis, bone diseases, arteriosclerosis, thrombosis, hypertension,
CC   indigestion, memory loss, obesity, hypoglycaemia, edema, insulin
CC   resistance, insulin allergy, fat metabolism disorders and cancer. This
CC   sequence represents a hamster GPR40 protein of the invention.
XX
SQ   Sequence 300 AA;

Query Match   81.8%; Score 1295; DB 7; Length 300;
Best Local Similarity   80.7%; Pred. No. 1.1e-120;
Matches 242; Conservative 21; Mismatches 37; Indels 0; Gaps 0;

Qy   1 MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLP 60
            ||||||||||||:||||:|||||||||||::|||:|||||||||||| ::||
Db   1 MALSPQLFFALYVSAFALGFPLNLLAIRGAVARARLRLTPNLVYTLHLACSDLLLAITLP 60

Qy   61 LKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRP 120
            |||||||||:|||| |||||||||||||||||||||||||||||||| |||||||||
Db   61 LKAVEALASGVWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAPPLGYQAFRRP 120

```
Db   61  VKAVEALASGAWPLPLPLCPVFVLVHPAPLYAGGGFLAALSAGRYLGAAFPFGYQAVRRP 120
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  121  CYSWGVCAAIWALVALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG 180
             |||||  ||||||||||   ::|||||||||||||||||||||||||||||||||||||
Db  121  RYSWGVCVAIWALVLCHMGLVLGLEAPGGWLNTTSSSLGINTPVNGSPVCLEAWDPNSAR 180
Qy  181  PARFSLSLLFFLPLAITAPCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY 240
             ||||||||||||||||||||||||||||||||  ||||||||||:||||
Db  181  PARLSFSILLFFVPLVITAPCYVGCLRALAHSGLSHKRKLRAAWAAGGAFLTLLLCLGPY 240
Qy  241  NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK 300
             ||||||::|:||||||||||:||||||||||||||||||||||||
Db  241  NASNVASFVNPDLGGSWRKLGLITGSWSVVLNPLVTGYLGASPGRGTVCTTRTQGGTIQK 300
```

RESULT 13
AAB82759

ID   AAB82759 standard; protein; 346 AA.
XX
AC   AAB82759;
XX
DT   29-OCT-2001 (first entry)
DE
DE   Human G-protein coupled receptor GPR 42.
XX
KW   GPR 42; G-protein coupled receptor 42; human; drug screening;
KW   dyslipidaemia; coronary heart disease; atherosclerosis; thrombosis;
KW   obesity; angina; kidney failure; peripheral vascular disease; stroke;
KW   diabetes; metabolic syndrome; syndrome X; antilipaemic; cardiant;
KW   antiatherosclerotic; thrombolytic; anorectic; antianginal; nephrotropic;
KW   vasotropic; cerebroprotective; antidiabetic; therapy.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   Domain   18..41
FT   /label= Transmembrane
FT   Domain   52..73
FT   /label= Transmembrane
FT   Domain   88..111
FT   /label= Transmembrane
FT   Domain   132..153
FT   /label= Transmembrane
FT   Domain   188..212
FT   /label= Transmembrane
FT   Domain   229..250
FT   /label= Transmembrane
FT   Domain   259..278
FT   /label= Transmembrane
XX
PN   WO200161359-A2.
XX
PD   23-AUG-2001.
XX
PF   19-FEB-2001; 2001WO-GB000684.
XX
PR   18-FEB-2000; 2000GB-00003900.
PR   22-MAR-2000; 2000GB-00007015.
XX
PA   (GLAX) GLAXO GROUP LTD.
XX
PI   Wise A, Brown AJ;
XX
DR   WPI; 2001-536581/59.
DR   N-PSDB; AAH26460.
XX
PT   Identification of an agent for the modulation of G-protein coupled
PT   receptor useful for the treatment of disease i.e. dyslipidemia or stoke.
XX
PS   Claim 7; Page 47-48; 53pp; English.
XX
CC   The present sequence is that of human G-protein coupled receptor 42 (GPR
CC   42). The invention is based on the finding that expression of GPR 42 and

---

CC   GPR 41 (see AAB82758) is restricted to adipose tissue. GPR 41 or GPR 42
CC   may therefore be used as a screening target for the identification and
CC   development of novel pharmaceutical agents for use in inhibiting
CC   lipolysis. Methods are claimed for identifying agents that modulate GPR
CC   41 or GPR 42 activity, involving: (1) contacting a test agent with GPR 41
CC   or GPR 42 or a variant polypeptide capable of coupling to a G-protein;
CC   and (2) monitoring for GPR 41 or GPR 42 activity in the presence of a G-
CC   protein, thereby determining whether the test agent modulates activity.
CC   Such agents, which may be an activator of GPR 41 or GPR 42, an inhibitor
CC   of lipolysis, or a polynucleotide encoding GPR 41, GPR 42 or variant
CC   polypeptide, are useful for the treatment of dyslipidaemia, coronary
CC   heart disease, atherosclerosis, thrombosis or obesity, angina, chronic
CC   renal failure, peripheral vascular disease, stroke, type II diabetes or
CC   metabolic syndrome (syndrome X) (all claimed)
XX
SQ   Sequence 346 AA;

Query Match   23.1%;   Score 366;   DB 4;   Length 346;
Best Local Similarity   34.1%;   Pred. No. 6.8e-28;
Matches   101;   Conservative   38;   Mismatches   129;   Indels   28;   Gaps   8;

```
Qy    1  MDLPPQLS--------FGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCS  51
            |||      ||     : | :|  ||:||:|
Db    1  MDTGPDQSYFSGNHWFVFSVVLLTPLVGLPLNLLALVVFVGKLRCRPVAVDVLLNLTAS  60
Qy   52  DLLLTVSLPLKAVEALASGAWPLPASLCPVPAVAHFPPLYAGGGFLAALSAGRYLGAAFP  111
            ||||||||||||||||:||| |:|  | :|||:||||||:|:|||:||
Db   61  DLLLLLFLPFRMVEAANGMHWPLPFIICPLSGFIPFTTIYLTALFLAAVSIERFLSVAHP  120
Qy  112  LGYQAFRRPCYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVC-  170
            ||||  |  |||:||  :||:: :|| |||| |:|
Db  121  LWYKTRPRLGQAGLVSVACWLLASAHCSVVVVIEFSGD-ISHSQGTNG--------TCY  170
Qy  171  LRAWDP--ASAGPARFSLSLLLFFLPLAITAPCYVGCLRALARSGLTHRRKLRAAWVAGG  228
            ||||    :: ||  |||:|||  |: ||| :|:|||:||||:::|||::|||
Db  171  LEFWKDQLAILLPVRLEMAVVLFVVPLIITSYCYSRLVWILGRGG-SHRRQRRVAGLVAA  229
Qy  229  ALLTLLLCVGPYNASNVASFLYPNLGGS--WRKLGLITGAWSVVLNPLVTGYLGRG  282
            || ||||||||  |:||||||||:||||  |||||||||:||||||||||:|||
Db  230  TLLNPLVCFGPYNVSHVVGYI---CGESPVRIYVTLLSTLNSCVDPFVYYSSSSG  282
```

RESULT 14
AAY90680
ID   AAY90680 standard; protein; 346 AA.
XX
AC   AAY90680;
XX
DT   21-AUG-2000 (first entry)
DE
DE   Human G protein-coupled receptor GPR41.
XX
KW   G protein-coupled receptor; GPCR; constitutively active;
KW   intracellular loop 3; transmembrane domain 6; drug screening; agonist;
KW   antagonist.
XX
OS   Homo sapiens.
XX
PN   WO200022129-A1.
XX
PD   20-APR-2000.
XX
PF   12-OCT-1999;   99WO-US023938.
XX
PR   13-OCT-1998;   98US-00170496.
XX
PA   (AREN-) ARENA PHARM INC.
XX
PI   Behan DP, Chalmers DT, Liaw CW;
XX
DR   WPI; 2000-329165/28.
DR   N-PSDB; AAA30762.
XX

PT   Non-endogenous constitutively activated human G protein-coupled
PT   receptors, useful for identifying agonists for use as pharmaceutical
PT   agents.
XX
PS   Example 1; Page 310-311; 341pp; English.
XX
CC   The invention relates to constitutively active, non-endogenous versions
CC   of endogenous human orphan G protein-coupled receptors (GPCRs, AAY90643-
CC   AAY90677 and AAY90683-AAY90687), and to DNA encoding them (AAA30709-A30743
CC   and AAA30775-A30779). The mutant proteins of the invention contain a
CC   mutation in a portion of the protein comprising intracellular loop 3
CC   (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, X,
CC   is substituted for an endogenous residue in IC3 at a position 16 amino
CC   acids N-terminal of an endogenous proline in TM6 to form a sequence X-
CC   (AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg or
CC   Ala, and is preferably Lys. When the endogenous residue at this position
CC   is Lys, this residue is replaced by His, Arg or preferably Ala. The 15
CC   amino acid stretch between the substituted amino acid and the Pro may be
CC   endogenous, non-endogenous, or a mixture of endogenous and non-endogenous
CC   residues. The constitutively active GPCRs are useful for identifying
CC   antagonists, agonists and partial agonists for use as pharmaceutical
CC   agents. The mutant proteins are also useful in research settings for
CC   elucidating the roles of the receptors in normal and diseased conditions.
CC   Antagonists for a particular GPCR are useful for treating diseases and
CC   disorders associated with that receptor. Because the novel mutant GPCRs
CC   are constitutively active, they can be used directly for screening of
CC   compounds without the need for endogenous ligands. The present sequence
CC   represents a human wild-type GPCR referred to in an exemplification of
CC   the invention
XX
SQ   Sequence 346 AA;

Query Match   21.9%; Score 347; DB 3; Length 346;
Best Local Similarity   32.3%; Pred. No. 5.4e-26;
Matches   95; Conservative   41; Mismatches 134; Indels   24; Gaps   6;

Qy   1 MDLPPQLS--------FGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCS 51
                       : :|: :|| |||||::
Db   1 MDTGPDQSYFSGNHWFVFSVYLLTFLVGLPLNLLALVVFVGKLQRRPVAVDVLLNLTAS 60

Qy   52 DLLLTVSLPLKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFP 111
                       |||| :|| || ||: :|:| |
Db   61 DLLLLLFLPFRMVEAANGMHWPLPFILCPLSGFIFFTTIYLTALFLAAVSIERFLSVAHP 120

Qy   112 LGYQAFRRPCYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCL 171
                |:| |:|:| :: |
Db   121 LWYKTRPRLGQAGLVSVACWLLASAHCSVVYVIEFSGD-ISHSQGTNG-------TCY 170

Qy   172 EAW---DPASAGPARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGG 228
                : || :||:| | :||:   : |||| |:|| |
Db   171 LEFRKDQLAILLPVRLEMAVVLFVVPLIITSYCYSRLVWILGRGG-SHRRQRRVAGLLAA 229

Qy   229 ALITLLLCVGPYNASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRG 282
                :|| ||||| ::||:
Db   230 TLLNFLVCFGPYNVSHVVVGYICGB-SPAWRIYVTLLSTLNSCVDPFVYFSSSG 282

RESULT 15
AAB82758
ID   AAB82758 standard; protein; 346 AA.
XX
AC   AAB82758;
XX
DT   29-OCT-2001 (first entry)
XX
DE   Human G-protein coupled receptor GPR 41.
XX
KW   GPR 41; G-protein coupled receptor 41; human; drug screening;
KW   dyslipidaemia; coronary heart disease; atherosclerosis; thrombosis;
KW   obesity; angina; kidney failure; peripheral vascular disease; stroke;
KW   diabetes; metabolic syndrome; syndrome X; antilipaemic; cardiant;
KW   antiatherosclerotic; thrombolytic; anorectic; antianginal; nephrotropic;
KW   vasotropic; cerebroprotective; antidiabetic; therapy.

XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   Domain   18..41
FT   /label= Transmembrane
FT   Domain   52..73
FT   /label= Transmembrane
FT   Domain   88..111
FT   /label= Transmembrane
FT   Domain   132..153
FT   /label= Transmembrane
FT   Domain   188..212
FT   /label= Transmembrane
FT   Domain   229..250
FT   /label= Transmembrane
FT   Domain   259..278
FT   /label= Transmembrane
XX
PN   WO200161359-A2.
XX
PD   23-AUG-2001.
XX
PF   19-FEB-2001; 2001WO-GB000684.
XX
PR   18-FEB-2000; 2000GB-00003900.
PR   22-MAR-2000; 2000GB-00007015.
XX
PA   (GLAX) GLAXO GROUP LTD.
XX
PI   Wise A, Brown AJ;
XX
DR   WPI; 2001-535581/59.
DR   N-PSDB; AAH26459.
XX
PT   Identification of an agent for the modulation of G-protein coupled
PT   receptor useful for the treatment of disease i.e. dyslipidemia or stoke.
XX
PS   Claim 7; Page 44-46; 53pp; English.
XX
CC   The present sequence is that of human G-protein coupled receptor 41 (GPR
CC   41). The invention is based on the finding that expression of GPR 41 and
CC   GPR 42 (see AAB82759) is restricted to adipose tissue. GPR 41 or GPR 42
CC   may therefore be used as a screening target for the identification and
CC   development of novel pharmaceutical agents for use in inhibiting
CC   lipolysis. Methods are claimed for identifying agents that modulate GPR
CC   41 or GPR 42 activity, involving: (1) contacting a test agent with GPR 41
CC   or GPR 42 or a variant polypeptide capable of coupling to a G-protein;
CC   and (2) monitoring for GPR 41 or GPR 42 activity in the presence of a G-
CC   protein, thereby determining whether the test agent modulates activity.
CC   Such agents, which may be an activator of GPR 41 or GPR 42, an inhibitor
CC   of lipolysis, or a polynucleotide encoding GPR 41, GPR 42 or variant
CC   polypeptide, are useful for the treatment of dyslipidaemia, coronary
CC   heart disease, atherosclerosis, thrombosis or obesity, angina, chronic
CC   renal failure, peripheral vascular disease, stroke, type II diabetes or
CC   metabolic syndrome (syndrome X) (all claimed)
XX
SQ   Sequence 346 AA;

Query Match   21.9%; Score 347; DB 4; Length 346;
Best Local Similarity   32.3%; Pred. No. 5.4e-26;
Matches   95; Conservative   41; Mismatches 134; Indels   24; Gaps   6;

Qy   1 MDLPPQLS--------FGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCS 51
                       : :|: :|| |||||::
Db   1 MDTGPDQSYFSGNHWFVFSVYLLTFLVGLPLNLLALVVFVGKLQRRPVAVDVLLNLTAS 60

Qy   52 DLLLTVSLPLKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFP 111
                       |||| :|| || ||: :|:| |
Db   61 DLLLLLFLPFRMVEAANGMHWPLPFILCPLSGFIFFTTIYLTALFLAAVSIERFLSVAHP 120

Qy   112 LGYQAFRRPCYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCL 171
                |:| |:|:| :: |

```
Db  121 LWYKTRPRLGQAGLVSVACWLLASAHCSVVYVIEFSGD-ISHSQGTNG---------TCY 170
                           | |   :::||:||   | ||:::|| : | | | | :
Qy  172 EAW--DPASAGPARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGG 228
            :       |--| |  ||::|| : | |-| :||:|| | :
Db  171 LEFRKDQLAILLPVRLEMAVVLFVVPLIITSYCYSRLVWILGRGG-SHRQRRVAGLLAA 229

Qy  229 ALLTLLLCVGPYNASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRG 282
            || |:|:||||| |:| :   :|| | : :|  |
Db  230 TLLNFLVCFGPYNVSHVVGYICGE-SPAWRIYVTLLSTLNSCVDPFVYFSSSG 282
```

Search completed: April 29, 2004, 12:10:03
Job time : 61 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2004, 12:07:54 ; Search time 20 Seconds
(without alignments)
1442.873 Million cell updates/sec

Title: US-10-202-687-2
Perfect score: 1584
Sequence: 1 MDLPPQLSFGLYVAAFALGF.........RGPGLKTVCAARTQGGKSQK 300

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match	Length	DB	ID	Description
1	1584	100.0	300	300	2	JC5714	G protein-coupled
2	366	23.1	346	346	2	JC5716	G protein-coupled
3	347	21.9	346	346	2	JC5715	G protein-coupled
4	332	21.0	330	330	2	JC5717	G protein-coupled
5	244	15.4	420	420	2	I51667	thrombin receptor
6	228	14.4	425	425	2	A37912	thrombin receptor
7	213	13.4	427	427	2	S17148	alpha-thrombin rec
8	205.5	13.0	363	363	2	I57940	somatostatin recep
9	204.5	12.9	355	355	2	JQ1231	interleukin-8 rece
10	204	12.9	432	432	2	A43448	thrombin receptor
11	201	12.7	358	358	2	A53752	interleukin-8 rece
12	198.5	12.5	328	328	2	JC4800	P2Y6 receptor - hu
13	197.5	12.5	328	328	2	I55450	G protein-coupled
14	195.5	12.3	362	362	2	S68207	G protein-coupled
15	194.5	12.3	350	350	2	A39445	interleukin-8 rece
16	194	12.2	363	363	2	I57955	somatostatin recep
17	194	12.2	364	364	2	JN0763	somatostatin recep
18	190.5	12.0	354	354	2	I53033	G protein-coupled
19	190.5	12.0	362	362	2	A57641	G protein-coupled
20	189.5	12.0	359	359	2	JC5277	G protein-coupled
21	188.5	11.9	352	352	1	S27357	complement C5a ana
22	186.5	11.8	362	362	2	B57641	G protein-coupled
23	186	11.7	375	375	2	A54946	P-2U nucleotide re
24	184	11.6	365	365	2	S68679	G protein-coupled
25	182.5	11.5	355	355	2	I49939	macrophage inflamm
26	182.5	11.5	360	360	2	A53611	interleukin-8 rece
27	180.5	11.4	355	355	2	A45177	chemokine (C-C) re
28	179	11.3	373	373	2	A47556	ATP receptor P2u -
29	179	11.3	384	384	2	A47249	brain-specific som
30	178	11.2	361	361	2	B45680	G protein-coupled
31	178	11.2	361	361	2	JC5653	G protein-coupled
32	175.5	11.1	388	388	2	JN0605	somatostatin recep
33	173.5	11.0	367	367	2	I49022	kappa opioid recep
34	173.5	11.0	367	367	2	JC2421	opioid receptor ho
35	173.5	11.0	367	367	2	I56520	G protein-coupled
36	173	10.9	365	365	2	S68208	G protein-coupled
37	172.5	10.9	308	308	2	I50241	G protein-coupled
38	172	10.9	352	352	2	S60024	bradykinin B1 rece
39	172	10.9	370	370	2	JC5549	heptahelical P2Y5-
40	171.5	10.8	356	356	2	S42096	interleukin-8 rece
41	171.5	10.8	372	372	2	B48227	delta opioid recep
42	171.5	10.8	374	374	2	S32785	G protein-coupled
43	171	10.8	384	384	2	JC4629	somatostatin recep
44	170.5	10.8	360	360	2	G02064	G protein-coupled
45	169.5	10.7	344	344	2	T09508	intron 17 purinerg

ALIGNMENTS

RESULT 1
JC5714
G protein-coupled receptor 40 - human
C;Species: Homo sapiens (man)
C;Date: 03-Dec-1997 #sequence_revision 23-Jan-1998 #text_change 21-Jul-2000
C;Accession: JC5714
R;Sawzdargo, M.; George, S.R.; Nguyen, T.; Xu, S.; Kolakowski Jr., L.F.; O'Dowd, B.F.
Biochem. Biophys. Res. Commun. 239, 543-547, 1997
A;Title: A cluster of four novel human G protein-coupled receptor genes occurring in clos
A;Reference number: JC5714; MUID:98008875; PMID:9344866
A;Accession: JC5714
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-300 <SAW>
A;Cross-references: GB:AF024687; NID:g2612945; PIDN:AAB86710.1; PID:g2612946
C;Superfamily: G protein-coupled receptor 43
C;Keywords: glycoprotein; lipoprotein; thiolester bond
F;13-32/Domain: transmembrane #status predicted <TM1>
F;43-64/Domain: transmembrane #status predicted <TM2>
F;83-102/Domain: transmembrane #status predicted <TM3>
F;125-144/Domain: transmembrane #status predicted <TM4>
F;184-206/Domain: transmembrane #status predicted <TM5>
F;222-243/Domain: transmembrane #status predicted <TM6>
F;260-280/Domain: transmembrane #status predicted <TM7>
F;155,165/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;43-64/Binding site: palmitate (Cys) (covalent) #status predicted
F;289/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match   100.0%;   Score 1584;   DB 2;   Length 300;
Best Local Similarity   100.0%;   Pred. No. 2.8e-125;
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLP 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLP 60

Qy   61 LKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRP 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRP 120

Qy  121 CYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 CYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG 180

Qy  181 PARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLCVGPY 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 PARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLCVGPY 240

Qy  241 NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK 300
```

Left column — (RESULT 2 continued)

```
Query Match             23.1%;  Score 366;  DB 2;  Length 346;
Best Local Similarity   34.1%;  Pred. No. 3e-23;
Matches 101; Conservative  38;  Mismatches 129;  Indels  28;  Gaps   8;

Qy    1 MDLPPQLS--------FGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCS 51
Db    1 MDTGPDQSYFSGNHWFVFSVYLLTFLVGLPLNLLALVVFVGKLRCRPVAVDVLLLNLTAS 60

Qy   52 DLLLTVSLPLKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGFLAALSAGRYLGAAFP 111
Db   61 DLLLLLFLPFRMVEAANGMHWPLPFILCPLSGFIFFTIYLTALFLAAVSIERFLSVAHP 120

Qy  112 LGYQAFRRPCYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVC- 170
Db  121 LWYKTRPRLGQAGLVSVACWLLASAHCSVVYVIEFSGD-ISHSQGTNG--------TCY 170

Qy  171 LEAWDP--ASAGPARFSLSLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGG 228
Db  171 LEFWKDQLAILLPVRLEMAVVLFVVPLIITSYCYSRLVWILGRGG-SHRRQRRVAGLVAA 229

Qy  229 ALLTLLLCVGPYNASNVASFLYPNLGGS--WRKLGLITGAWSVVLNPLVTGYLGRG 282
Db  230 TLLNFLVCFGPYNVSHVVGYI---CGESPVWRIYVTLLSTLNSCVDPFVYYSSSG 282
```

RESULT 3
JC5715
G protein-coupled receptor 41 - human
C;Species: Homo sapiens (man)
C;Date: 03-Dec-1997 #sequence_revision 29-Sep-1999
C;Accession: JC5715
R;Sawzdargo, M.; George, S.R.; Nguyen, T.; Xu, S.; Kolakowski Jr., L.F.; O'Dowd, B.F.
Biochem. Biophys. Res. Commun. 239, 543-547, 1997
A;Title: A cluster of four novel human G protein-coupled receptor genes occurring in clo
A;Reference number: JC5714; MUID:98008875; PMID:9344866
A;Accession: JC5715
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-346 <SAW>
A;Cross-references: GB:AF024688; NID:g2612947; PIDN:AAB86711.1; PID:g2612948
C;Superfamily: G protein-coupled receptor 43
C;Keywords: glycoprotein; lipoprotein; thiolester bond
F;18-41/Domain: transmembrane #status predicted <TM1>
F;52-73/Domain: transmembrane #status predicted <TM2>
F;90-111/Domain: transmembrane #status predicted <TM3>
F;132-153/Domain: transmembrane #status predicted <TM4>
F;188-212/Domain: transmembrane #status predicted <TM5>

Right column — (RESULT 3 continued)

F;229-250/Domain: transmembrane #status predicted <TM6>
F;259-278/Domain: transmembrane #status predicted <TM7>
F;166/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;295/Binding site: palmitate (Cys) (covalent) #status predicted

```
Query Match             21.9%;  Score 347;  DB 2;  Length 346;
Best Local Similarity   32.3%;  Pred. No. 1.2e-21;
Matches  95;  Conservative  41;  Mismatches 134;  Indels  24;  Gaps   6;

Qy    1 MDLPPQLS--------FGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCS 51
Db    1 MDTGPDQSYFSGNHWFVFSVYLLTFLVGLPLNLLALVVFVGKLQRRPVAVDVLLLNLTAS 60

Qy   52 DLLLTVSLPLKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGFLAALSAGRYLGAAFP 111
Db   61 DLLLLLFLPFRMVEAANGMHWPLPFILCPLSGFIFFTIYLTALFLAAVSIERFLSVAHP 120

Qy  112 LGYQAFRRPCYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCL 171
Db  121 LWYKTRPRLGQAGLVSVACWLLASAHCSVVYVIEFSGD-ISHSQGTNG--------TCY 170

Qy  172 EAW--DPASAGPARFSLSLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGG 228
Db  171 LEFRKDQLAILLPVRLEMAVVLFVVPLIITSYCYSRLVWILGRGG-SHRRQRRVAGLLAA 229

Qy  229 ALLTLLLCVGPYNASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRG 282
Db  230 TLLNFLVCFGPYNVSHVVGYICGE-SPAWRIYVTLLSTLNSCVDPFVYYSSSG 282
```

RESULT 4
JC5717
G protein-coupled receptor 43 - human
C;Species: Homo sapiens (man)
C;Date: 03-Dec-1997 #sequence_revision 29-Sep-1999
C;Accession: JC5717
R;Sawzdargo, M.; George, S.R.; Nguyen, T.; Xu, S.; Kolakowski Jr., L.F.; O'Dowd, B.F.
Biochem. Biophys. Res. Commun. 239, 543-547, 1997
A;Title: A cluster of four novel human G protein-coupled receptor genes occurring in clo
A;Reference number: JC5714; MUID:98008875; PMID:9344866
A;Accession: JC5717
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-330 <SAW>
A;Cross-references: GB:AF024690; NID:g2612951; PIDN:AAB86713.1; PID:g2612952
C;Superfamily: G protein-coupled receptor 43
C;Keywords: glycoprotein
F;15-34/Domain: transmembrane #status predicted <TM1>
F;46-67/Domain: transmembrane #status predicted <TM2>
F;86-105/Domain: transmembrane #status predicted <TM3>
F;128-147/Domain: transmembrane #status predicted <TM4>
F;181-204/Domain: transmembrane #status predicted <TM5>
F;220-242/Domain: transmembrane #status predicted <TM6>
F;256-276/Domain: transmembrane #status predicted <TM7>
F;151,167/Binding site: carbohydrate (Asn) (covalent) #status predicted

```
Query Match             21.0%;  Score 332;  DB 2;  Length 330;
Best Local Similarity   31.0%;  Pred. No. 2e-20;
Matches  93;  Conservative  53;  Mismatches 120;  Indels  34;  Gaps  12;

Qy   12 YVAAFALGFPLNVLAIRGATAHARL-RLTPSLVYALNLGCSDLLLTVSLPLKAVEALASG 70
Db   14 YIIIFLTGLPANLLALRAFVGRIRQPQPAPVHILLSLTLADLLLLLLLPFKIIEAASNF 73

Qy   71 AWPLPASLCPVFAVAHFFPLYAGGFLAALSAGRYLGAAFPLGYQAFRRPCYSWGVCAAI 130
Db   74 RWYLPKVVCALTSFGYSSIYCSTWLLAGISIERYLGVAFPVQYKLSRRPLY--GVIAAL 131

Qy  131 --WALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPV-CLEAWDPASAG---PARF 184
Db  132 VAWVMSFGHCTIVIIVQ-------YLNTTEQVR---SGNEITCYENFTDNQLDVVLPVRL 181

Qy  185 SLSLLFFLPLAITAFCY---VGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPYN 241
```

```
Matches     65;  Conservative   46;  Mismatches  104;  Indels   42;  Gaps    8;

Qy    11  LYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLPLKAVEALASG   70
              :  ::|||:|:|  |||:||:|:|::|
Db   108  VYTGVFVVSLPLNIMAIVVFILKMKVK-KPAVVYMLHLATADVLFVSVLPFKISYYFSGS  166

Qy    71  AWPLPASLCPVPAVAHFFPLYAGGFLAALSAGRYLGAAFPLGYQARRPCYSWGVCAAI   130
              :  ::|   | :||:|  ::||| |
Db   167  DWQFGSBLCRFVTAAFYCNMYASILLMTVISIDRFLAVVYPMQSLSWRTLGRASFTCLAI  226

Qy   131  WALVLCHLGLV------FGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPA-------  177
           |||   || :
Db   227  WALAIA--GVVPLVLKEQTIQVPG---------LNITT-------CHDVLNETLLEGYA   268

Qy   178  ---SAGPARFSLSLLFFLPLAITAFCYVGCLRALARSGLTHR-RKLRAAWVAGGALLTL   233
                :  :||: ::|:|  | : :||   ::
Db   269  YYFSA------FSAVFFVPLLISTVCYVSIIRCLSSSAVANRSKKSRALFLSAAVPCIF   322

Qy   234  LLCVGPYNASNVASFLY                                            250
           ::|||   :  ::   :
Db   323  IICFGPTNVLLIAHYSF                                            339


RESULT 7
S17148
alpha-thrombin receptor - Chinese hamster
C;Species: Cricetulus griseus (Chinese hamster)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-Nov-1999
C;Accession: S17148
R;Rasmussen, U.B.; Vouret-Craviari, V.; Jallat, S.; Schlesinger, Y.; Pages, G.; Pavirani,
  FEBS Lett. 288, 123-128, 1991
A;Title: cDNA cloning and expression of a hamster alpha-thrombin receptor coupled to Ca(2
A;Reference number: S17148; MUID:91348247; PMID:1652467
A;Accession: S17148
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-427 <RAS>
A;Cross-references: EMBL:X61958; NID:g940495; PIDN:CAA43957.1; PID:g49538
C;Keywords: G protein-coupled receptor; transmembrane protein

Query Match           13.4%;  Score 213;  DB 2;  Length 427;
Best Local Similarity 23.4%;  Pred. No. 2.4e-10;
Matches     68;  Conservative   53;  Mismatches  122;  Indels   48;  Gaps    9;

Qy    11  LYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLPLKAVEALASG   70
              :  ::|||:|:|  |||:||:|:|::|
Db   110  VYTFVFVVSLPLNILAIAVFVLKMCVK-KPAVVYMLHLAMADVLFVSVLPLKISYYFSGS  168

Qy    71  AWPLPASLCPVPAVAHFFPLYAGGFLAALSAGRYLGAAFPLGYQARRPCYSWGVCAAI   130
              :  ::|   | :||:|  ::||| |
Db   169  DWQFGSGMCRFATAAFYCNMYASIMLMTVISIDRFLAVVYPIQSLSWRTLGRANFTCLVI  228

Qy   131  WALVLCHLGLVFGL------EAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPA-------  177
           | :   || :
Db   229  WVMAI--MGVVPLLLKEQTTRVPG---------LNITT-------CHDVLNETLLQGFYS  270

Qy   178  ---SAGPARFSLSLLFFLPLAITAFCYVGCLRALARSGLTHR-RKLRAAWVAGGALLTL   233
                :  :||: ::|:|  | : :||   ::
Db   271  YYFSA------FSAVFFLVPLIISTICYMSIIRCLSSSVANRSKKSRALFLSAAVPCVF   324

Qy   234  LLCVGPYNASNVASFLYPNLGGSWRK------LGLITGAWSVVLNPLVTGY          278
           ::|||   :  ::   :
Db   325  IVCFGPTNVLLIMHYLLLSDSPATEKAYFAYLLCVCVSSVSCCIDPLIYYY          375


RESULT 8
I57940
somatostatin receptor 5 - rat
N;Alternate names: somatotropin release-inhibiting factor subtype 28 receptor
C;Species: Rattus norvegicus (Norway rat)
C;Date: 02-Aug-1996 #sequence revision 02-Aug-1996 #text_change 24-Nov-1999
C;Accession: I57940; I57949; S39244
R;O'Carroll, A.M.; Lolait, S.J.; Konig, M.; Mahan, L.C.
```

```
Db   182  ELCLVLFFIPMAVTIFCYWRFVWIMLSQPLVGAQRRR--RAVGLAVVTLNLPLVCFGPYN  239
             |:||| |:|:|:|  ||   |  ||      |  |   |  :|
Qy   242  ASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGR------GPGLKTVCAARTQG  295
             ::  |  |       | :      ||      | :|:|
Db   240  VSHLVGY-HQRKSPWRSIAVVFSSLNASLDPLLFYFSSSSVVRRAFGRGLQVL---RNQG  295


RESULT 5
I51667
thrombin receptor - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C;Accession: I51667
R;Gerszten, R.E.; Chen, J.; Ishii, M.; Ishii, K.; Nanevicz, T.; Turck, C.W.; Vu, T.H.; C
  Nature 368, 648-651, 1994
A;Title: Thrombin receptor's specificity for agonist peptide is determined by its extrac
A;Reference number: I51667; MUID:9419529; PMID:8145852
A;Accession: I51667
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-420 <GER>
A;Cross-references: EMBL:U09632; NID:g495197; PIDN:AAA18498.1; PID:g495198

Query Match           15.4%;  Score 244;  DB 2;  Length 420;
Best Local Similarity 25.0%;  Pred. No. 6e-13;
Matches     72;  Conservative   44;  Mismatches  126;  Indels   46;  Gaps    7;

Qy    11  LYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLPLKAVEALASG   70
           |||   | ||||:|||  :||:|:|
Db   107  LYTVVFIVGLPLNLLAIIIFLFKMKVR-KPAVVYMLNLAIADVFFVSVLPFKIAYHLSGN  165

Qy    71  AWPLPASLCPVPAVAHFFPLYAGGFLAALSAGRYLGAAFPLGYQARRPCYSWGVCAAI   130
              :  ::|   |   ||:|  : ||| |
Db   166  DWLFGPGMCRIVTAIFYCNMYCSVLLIASISVDRFLAVVYPMHSLSWRTMSRAYMACSFI  225

Qy   131  WALVLCHLGLIVFGLEAPGGWLDHSNTSLGINTPVNGSP------VCLEAWDPASAGPARF 184
           | : :
Db   226  WLISIA----------------STIPLLVTBQTQKIPRLDITTCHDVLDLKDLKDFYIY  268

Qy   185  --SLSLLLFFLPLAITAFCYVGCLRALARSGLTHR-RKLRAAWVAGGALLTLLLCVGPY   240
              |:| ::|:|  | : :||   ::
Db   269  YFSSFPCLLFFVPFIITTICYIGIIRSLSSSIENSCKKTRALFLAVVLCVFIICFGPT   328

Qy   241  N-----ASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGY                 278
                  ::
Db   329  NVLFITHYIQEANEFLYPAY--------ILSACVGSVSVSCCLDPLIYYY          368


RESULT 6
A37912
thrombin receptor precursor - human
C;Species: Homo sapiens (man)
C;Date: 22-Jan-1993 #sequence_revision 22-Jan-1993 #text_change 05-Nov-1999
R;Vu, T.K.H.; Hung, D.T.; Wheaton, V.I.; Coughlin, S.R.
  Cell 64, 1057-1068, 1991
A;Title: Molecular cloning of a functional thrombin receptor reveals a novel proteolytic
A;Reference number: A37912; MUID:91168254; PMID:1672265
A;Accession: A37912
A;Molecule type: mRNA
A;Residues: 1-425 <VUA>
A;Cross-references: GB:M62424; NID:g339676; MUID:91168254; PIDN:AAA36743.1; PID:g339677
C;Genetics:
A;Gene: GDB:F2R
A;Cross-references: GDB:127737; OMIM:187930
A;Map position: 5q13-5q13
C;Keywords: G protein-coupled receptor; glycoprotein; transmembrane protein
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-425/Product: thrombin receptor #status predicted <MAT>

Query Match           14.4%;  Score 228;  DB 2;  Length 425;
Best Local Similarity 25.3%;  Pred. No. 1.3e-11;
```

Mol. Pharmacol. 42, 939-946, 1992
A;Title: Molecular cloning and expression of a pituitary somatostatin receptor with pref
A;Reference number: I57940; MUID:93125499; PMID:1362243
A;Accession: I57940
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-363 <OCA1>
A;Cross-references: GB:L04535; NID:g409238; PIDN:AAA17029.1; PID:g409239
R;O'Carroll, A.M.; Lolait, S.J.; König, M.; Mahan, L.C.
Mol. Pharmacol. 44, 1278, 1993
A;Title: Molecular cloning and expression of a pituitary somatostatin receptor with pref
A;Reference number: I57949; MUID:94088493; PMID:8264565
A;Accession: I57949
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 341-363 <OCA2>
A;Cross-references: GB:S67370; NID:g455947; PIDN:AAB29371.1; PID:g455948
A;Experimental source: pituitary
R;Penetta, R.; Greenwood, M.; Patel, Y.C.
submitted to the EMBL Data Library, August 1993
A;Description: Correction of the nucleotide and amino acid sequence of the rat somatosta
A;Reference number: S39244
A;Accession: S39244
A;Molecule type: mRNA
A;Residues: 309-363 <PEN>
A;Cross-references: EMBL:X74828; NID:g433911; PIDN:CAA52825.1; PID:g433912
C;Genetics:
A;Gene: SSTR5
C;Superfamily: vertebrate rhodopsin

Query Match   13.0%;   Score 205.5; DB 2;   Length 363;
Best Local Similarity   25.4%;   Pred. No. 8.7e-10;
Matches   78; Conservative   49; Mismatches 135; Indels   45; Gaps   12;

Qy   11 LYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLPLKAVEALASG 70
         ||| : || ||||||||||| :|: ||||||||||||||||||||||||||
Db   43 LYLLVCTVGLSGNTLVIYVLRHAKMK-TVTNVYILNLAVADVLFMLGLPFLATQNAVVS 101

Qy   71 AWPLPASLCPVF-----AVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRPCYSWGV 126
         : || | : ||||:||:|
Db 102 YWPFGSFLCRLVMTLDGINQFTSIFC----LMVMSVDRYLAVVHPLRSARWRPRVAKMA 157

Qy 127 CAAIWAL-VLCHLGLVFGLEAPGWLDHSNTSLGINTPVNGSPVCLEAW-DPASAGPARF 184
         | || | | :|| :
Db 158 SAAVWVFSILMSLPLLVFADVQEGW----------GTCNLSWPEPVGLWGAAF 200

Qy 185 --SLSLLFFLPLAITAFCYVGCLRALARSGL---THRRKLRAAWVAGGALLTLLLCVG- 238
           :| |: | | : |||| ||:|| ||||||: :|
Db 201 ITYTSVLGFFGPLLVICLCYLLIVVKVKAAGMRVGSSRRRSBEKVTRMVVVVLVFVGC 260

Qy 239 --PYNASNVASFLY------PNLGGSWRKLGLITGAWSVVLNPLVTGYLG---RGPGLKTV 288
           ||||:||: || ||| : |: ||:|
Db 261 WLPFFIVNIVNLAFTLPEBPTSAGLYFFVVVLSYANSCA-NPLLYGFLSDNFRQSFRKVL 319

Qy 289 CAARTQG 295
         | || |
Db 320 CLRRGYG 326

RESULT 9
JQ1231
interleukin-8 receptor - rabbit
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 05-Nov-1999
C;Accession: JQ1231; A46483
R;Beckmann, M.P.; Munger, W.E.; Kozlosky, C.; VandenBos, T.; Price, V.; Lyman, S.; Gerar
Biochem. Biophys. Res. Commun. 179, 784-789, 1991
A;Title: Molecular characterization of the interleukin-8 receptor.
A;Reference number: JQ1231; MUID:91378994; PMID:1898400
A;Accession: JQ1231
A;Molecule type: DNA
A;Residues: 1-355 <BEC>
A;Cross-references: GB:M74240; NID:g165438; PIDN:AAA31375.1; PID:g165439

R;Lee, J.; Kuang, W.J.; Rice, G.C.; Wood, W.I.
J. Immunol. 148, 1261-1264, 1992
A;Title: Characterization of complementary DNA clones encoding the rabbit IL-8 receptor.
A;Reference number: A46483; MUID:92148149; PMID:1737938
A;Accession: A46483
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-355 <LEE>
A;Cross-references: GB:M82873; NID:g165440; PIDN:AAA31376.1; PID:g165441
A;Experimental source: neutrophils
A;Note: sequence extracted from NCBI backbone (NCBIN:81526, NCBIP:81530)
C;Superfamily: vertebrate rhodopsin
C;Keywords: G protein-coupled receptor; transmembrane protein

Query Match   12.9%;   Score 204.5; DB 2;   Length 355;
Best Local Similarity   25.8%;   Pred. No. 1e-09;
Matches   83; Conservative   39; Mismatches 117; Indels   83; Gaps   12;

Qy   13 VAAFALGFPLNVLA---IRGATAHARLRLTPSLVYALNLGCSDLLLTVSLPLKAVEALAS 69
         : |:| | |: :: |||| ||||| :: |
Db   48 VVIYALVFLLSLLGNSLVMLVILYSRSNRSVTDVYLLNLAMADLLFALTMPIWAVS--KE 105

Qy   70 GAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRPCYSWGVCAA 129
          || : ||: | | : |::|:|| |||| :   : | | | :|
Db 106 KGWIFGTPLCKVVSLVKEVNFYSGILLLACISVDRYLAIVHATRTLTQKRHLVKF-ICLG 164

Qy 130 IWALVLCHLGLVFGLEAPGWLDHSNTSLGINTPVNGSPVCLEAWDPASAGPARFSLSLL 189
         |||::| | | |:   ::| :
Db 165 IWALSLI-LSLPFFLFRQ------VFSPNNSSPVC--YEDLGHNTAKWRMVLR 208

Qy 190 L------FFLPLAITAFCYVGCLRAL--ARSGLTHRRKLRAAWVAGGALLTLLLCVGPYN 241
                 :|:| | | |||| :|| : ||: |: :|:: :| | |||
Db 209 ILPHTFGFIILPLLVMLFCYGFTLRTLFQAHMGQKH---RAMRVIFAVVLIFLLCWLPYN 264

Qy 242 ---------ASNVASFLYPNLGGSWRKLGLGLITGAWSVVLN 272
                  |:|:|:| :   :| | |||
Db 265 LVLLADTLMRTHVIQETCQRRNDIDRALDATEILGFLHS--------CLN 306

Qy 273 PLVTGYLGRG--PGLKTVCAAR 292
         |::   :| ||:|| ||||
Db 307 PIIYAFIGQNFRNGFLKMLAAR 328

RESULT 10
A43448
thrombin receptor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C;Accession: A43448
R;Zhong, C.; Hayzer, D.J.; Corson, M.A.; Runge, M.S.
J. Biol. Chem. 267, 16975-16979, 1992
A;Title: Molecular cloning of the rat vascular smooth muscle thrombin receptor. Evidence
A;Reference number: A43448; MUID:92381002; PMID:1324917
A;Accession: A43448
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-432 <ZHO>
A;Cross-references: GB:M81642; NID:g207465; PIDN:AAA42274.1; PID:g207466
A;Experimental source: RASM aortic smooth muscle cells
A;Note: sequence extracted from NCBI backbone (NCBIN:111973, NCBIP:111974)
C;Keywords: G protein-coupled receptor; transmembrane protein

Query Match   12.9%;   Score 204; DB 2;   Length 432;
Best Local Similarity   24.5%;   Pred. No. 1.4e-09;
Matches   63; Conservative   46; Mismatches 104; Indels   44; Gaps   9;

Qy   11 LYVAAFALGFPLNVLAIRGATAHARLRL-TPSLVYALNLGCSDLLLTVSLPLKAVEALAS 69
         || : | |||:|| ::   : | |:: |:|| ||| |||:| |:: |
Db 115 VYTFVFIVSLPLNILAI--AVFVFRMKVKKPAVVMLHLAMADVLFVSVLPFKISYYFSG 172

Qy   70 GAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRPCYSWGVCAA 129
          :| : || | | : |: |: || ||: |: | |
Db 173 TDWQFGSGMCRFATAACYCNMYASIMLMTVISIDRFLAVVPIQSLSWRTLGRANFTCVV 232

Qy 130 IWALVLCHGLVFGL-------EAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPA------ 177
        ||·|||·|:|||:| ·|·   :|·   :|
Db 233 IWVMAI--MGVVPLLLKEQTTQVPG----------LNITT-------CHDVLNETLLHGFY 274

Qy 178 ---SAGPARFSLSLLFFLPLAITAFCYVGCLRALARSGLTHR-RKLRAAWVAGGALLT 232
          ||·|||·||||||·||: || ·||: |||| ·: ||·|| :::
Db 275 SYYFSA------FSAIFFLVPLIISTVCYTSIIRCLSSSAVANRSKKSRALFLSSAAVFCI 328

Qy 233 LLLCVGPYNASNVASFL 249
        ·:| ||·||| :||
Db 329 FIVCFGPTNVLLIVHYL 345

RESULT 11
A53752
interleukin-8 receptor (clone 5B1a) - rabbit
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 05-Nov-1999
C;Superfamily: vertebrate rhodopsin
C;Keywords: G protein-coupled receptor; transmembrane protein
R;Prado, G.N.; Thomas, K.M.; Suzuki, H.; LaRosa, G.J.; Wilkinson, N.; Folco, E.; Navarro
J. Biol. Chem. 269, 12391-12394, 1994
A;Title: Molecular characterization of a novel rabbit interleukin-8 receptor isotype.
A;Reference number: A53752; MUID:94230294; PMID:8175642
A;Accession: A53752
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-358 <PRA>
A;Cross-references: GB:L24445; NID:g437661; PIDN:AAA31378.1; PID:g437662
C;Superfamily: G protein-coupled receptor; transmembrane protein

Query Match   12.7%;   Score 201;   DB 2;   Length 358;
Best Local Similarity   25.5%;   Pred. No. 2e-09;
Matches   89;   Conservative 34;   Mismatches 122;   Indels 104;   Gaps 14;

Qy   2 DLPP---------QLSFGLYVA--APALGFPLNVLA--IRGATAHARLRLTPSLVYA 45
         |||| ·||||·|| ·||·|||||·| :|·|||·|||| :||
Db   26 DLPPTLLDSAPCRSESSLETNSYVVLITYILVFLLSLLGNSLVMLVILYSRSTCSVTDVL 85

Qy   46 LNLGCSDLLLTVSLPLKAVEALASGAWPLPASLCPVFAVAHFPPLYAGGGFLAALSAGRY 105
        |||·|·|||||·|·||·   ·| :|||·|·||·|| ||||| ||·|· :|||·|| |·|
Db   86 LNLAIADLLFATTLPIWAASKV--HGWTFGTPLCKVVSLVKEVNPYSGILLLACISVDRY 143

Qy 106 LGAAFPLGYQAPRRP-----CYS-WGVCAIWALVLCHLGLVFGLEAPGGWLDHSNTSLG 159
        || :| |·   ·||· | |·| ||||||·|||||||||||·||
Db 144 LAIVHATRTMIQKRHLVKFICLSWMGVSLILSLPILLFRNAIF------------------- 186

Qy 160 INTPVNGSPVCLEAWDPASAGPARPSLSLL----FFLPLAITAFCYVGCLRAL--ARSG 213
        ···· |·|| ·| ||·||| ||·|||||||||| ·|||| :|·
Db 187 ---PPNSSPVCYEDMGNSTA-KWRMVLRIILPQTFGPILPLLVMLFCYVFTLRTLFQAHMG 242

Qy 214 LTHRRKLRAAWVAGGALLTLLLCVGPYN------------------------ASN 244
        ·||||·| |||·|||·|||·||| |||| |||
Db 243 QKH----RAMRVIFAVVLIFLLCWLPYNLVLLTDTLMRTHVIQETCERRNDIDRALDATE 298

Qy 245 VASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGR--GPGLKTVCAA 291
        ·:·||· |·|| ·|· :· ··|·||·|||| || || ·|·| |·:
Db 299 ILGFLHS-------------------CLNPIIYAFIGQKFRYGLLKILAA 329

RESULT 12
JC4800
P2Y6 receptor - human
C;Species: Homo sapiens (man)
C;Date: 15-Oct-1995 #sequence_revision 16-Aug-1996 #text_change 17-Nov-2000
C;Accession: JC4800; G02514
R;Communi, D.; Parmentier, M.; Boeynaems, J.M.
Biochem. Biophys. Res. Commun. 222, 303-308, 1996
A;Title: Cloning, functional expression and tissue distribution of the human P2Y6 recept
A;Reference number: JC4800; MUID:96224498; PMID:8670200
A;Accession: JC4800
A;Molecule type: mRNA

A;Residues: 1-328 <COM>
A;Cross-references: EMBL:X97058
A;Experimental source: placenta
R;Hammet, F.; Southey, M.C.; Somers, G.R.; Hutchins, A.M.; Venter, D.J.
submitted to the EMBL Data Library, March 1996
A;Reference number: H01373
A;Accession: G02514
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 'M',4-328 <HAM>
A;Cross-references: EMBL:U52464; NID:g1407632; PIDN:AAB03572.1; PID:g1407633
C;Genetics:
A;Gene: P2Y6
C;Superfamily: ATP receptor P2u
C;Keywords: glycoprotein; placenta; receptor; transmembrane protein
F;26-52/Domain: transmembrane #status predicted <TMM1>
F;63-86/Domain: transmembrane #status predicted <TMM2>
F;104-122/Domain: transmembrane #status predicted <TMM3>
F;143-167/Domain: transmembrane #status predicted <TMM4>
F;193-216/Domain: transmembrane #status predicted <TMM5>
F;241-264/Domain: transmembrane #status predicted <TMM6>
F;283-305/Domain: transmembrane #status predicted <TMM7>
F;5,173/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match   12.5%;   Score 198.5;   DB 2;   Length 328;
Best Local Similarity   27.4%;   Pred. No. 3e-09;
Matches   81;   Conservative 29;   Mismatches 135;   Indels 51;   Gaps 12;

Qy   3 LPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLPLK 62
        ||| ·|||·|·|·|||||··||· |·|||·| ·||·||||||||·|||·||| ||
Db   29 LPP-----VYSAVLAAGLPLNICVIT-QICTSRRAJTRTAVTTLNLALADLLYACSLPLL 82

Qy   63 AVEALASGAWPLPASLCPVFAVAHFPPLYAGGGFLAALSAGRYLGAAFPLG----YQAFRR 119
        ||·|·||||·||·|·|||| ||||||||||||||||||·|||||||||| ·|||||
Db   83 IYNYAQGDHWPFGDFACRLVRFLFYANLHGSILFLTCISFQRYLGICHPLAPWHKRGGRR 142

Qy 120 PCYSWGVCAAIWALVL--CHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPA 177
        ·| |||| |·   ·| ||·||·|·|||||||||·||||||||||·||||·|||
Db 143 A--AWLVCVAWLAVTTQCLPTAIF---------AATGIQ--RNRTVCYDLSPPA 184

Qy 178 SAG--PARFSLSLLFFLPLAITAFCYVGCLRA--LARSG----LTHRRKLRAAWVAG 227
        ||· ||·||||·|||||||||||||||||· ||·|| ||||||·|||||||
Db 185 LATHYMPYGMALTVIGFLLPPAALLACY--CLLACRLCRQDGPAEPVAQERRGKAARMAV 242

Qy 228 GALLTLLLCVGPYNASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGP 283
        ·:||||||||||||||||||||||||||||||||||||||||||||||·||||||||
Db 243 YLAVRSTPGVPCTVLEAFAAAYKGTRP 288

Qy 243 VVAAAFAISFLPFHITKTA-------------YLAVRSTPGVPCTVLEAFAAAYKGTRP 288

RESULT 13
I55450
G protein-coupled p2 receptor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 19-May-2000
C;Accession: I55450
R;Chang, K.; Hanaoka, K.; Kumada, M.; Takuwa, Y.
J. Biol. Chem. 270, 26152-26158, 1995
A;Title: Molecular cloning and functional analysis of a novel P2 nucleotide receptor.
A;Reference number: I55450; MUID:96064682; PMID:7592819
A;Accession: I55450
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-328 <RES>
A;Cross-references: GB:D63665; NID:g1066007; PIDN:BAA09816.1; PID:g1066008
C;Superfamily: ATP receptor P2u
C;Keywords: G protein-coupled receptor

Query Match   12.5%;   Score 197.5;   DB 2;   Length 328;
Best Local Similarity   26.2%;   Pred. No. 3.7e-09;
Matches   77;   Conservative 30;   Mismatches 140;   Indels 47;   Gaps 10;

Qy   3 LPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLPLK 62
        ||| ·||| |·|·|||||··||· |·|||·| ·||·|||||||||·|||·||| ||

```
Db   29 LPP------VYSVVLVVGLPLNVCVIAQICA-SRRTLTRSAVYTLNLALADLLYACSLPLL 82

Qy   63 AVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAARPLG---YQAFRR 119

Db   83 IYNYARGDHWPFGDLACRLVRFLFYANLHGSILFLTCISFQRYLGICHPLAPWHKRGGRR 142

Qy  120 PCYSWGVCAAIWALVL--CHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPA 177

Db  143 A--AWVVCCGVVWLVVTAQCLPTAVF-----------------AATGIQ--RNRTVCYDLSPPI 184

Qy  178 SAG---PARFSLSLLFFLPLAITAFCYVGCLRALARSG-----LTHRRKLRAAWVAGGA 229

Db  185 LSTRYLPYGMALTVIGFLLPFTALLACYCRMARRLCRQDGPAGPVAQERRSKAARMAVVV 244

Qy  230 LLTILLCVGPYNASNVASPLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGP 283

Db  245 AAVFVISFLPFHITKTA----------YLAVRSTPGVSCPVLETPAAAYKGTRP 288
```

RESULT 14
S68207
G protein-coupled receptor 6C.1 - human
C;Species: Homo sapiens (man)
C;Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 29-Sep-1999
C;Accession: S68207
R;An, S.; Tsai, C.; Goetzl, E.J.
FEBS Lett. 375, 121-124, 1995
A;Title: Cloning, sequencing and tissue distribution of two related G protein-coupled re
A;Reference number: S68207; MUID:96087098; PMID:7498459
A;Accession: S68207
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-362 <ANS>
A;Cross-references: EMBL:U35399; NID:g1015399; PIDN:AAA79060.1; PID:g1015420; PIDN:AAA79061.1; PID:g1015421
C;Superfamily: G protein-coupled receptor 4
C;Keywords: G protein-coupled receptor

Query Match   12.3%;   Score 195.5; DB 2;   Length 362;
Best Local Similarity   27.2%;   Pred. No. 5.9e-09;
Matches 74; Conservative 30; Mismatches 115; Indels 53; Gaps 10;

```
Qy    4 PPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRITPSLVYALNLGCSDLLLTVSLPLKA 63

Db   20 PP---SLYIFVIGVGLPTNCLALWAAYRQVQR-NELGVYLNLSIADLLYICSLPLWV 74

Qy   64 VEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRPCYS 123

Db   75 DYFLHHDNWIHGPGSCKLFPGFIFYTNIYISIAFLCCISVDRYLAVAHPLRFARLRRVKTA 134

Qy  124 WGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPV-------------- 169

Db  135 VAVSSVVWA----------------TELG----ANSAPLFHDELFRDRYNHTF 167

Qy  170 CLEAWDPASAGPARPFSLSLLL--FFLPLAITAFCYVGCLRALARSGLTHRR-KLRAAWVA 226

Db  168 CFEKF--PMEGWVAWMNLYRVFVGFLFPWALMLLSYRGILRAVRGSVSTERQBKAKIKRLA 226

Qy  227 GGALLTLLLCVGPYNASNVA-SFLYPNLGGSW 257

Db  227 LSLIAIVLVCFAPYHVLLLSRSAIY--LGRPW 256
```

RESULT 15
A39445
interleukin-8 receptor type A - human
N;Alternate names: interleukin-8 receptor, high-affinity
C;Species: Homo sapiens (man)
C;Date: 22-Jan-1993 #sequence_revision 12-Apr-1996 #text_change 05-Nov-1999
C;Accession: I37449; I38710; I38711; A39445
R;Mollereau, C.; Muscatelli, F.; Mattei, M.G.; Vassart, G.; Parmentier, M.
Genomics 16, 248-251, 1993
A;Title: The high-affinity interleukin 8 receptor gene (IL8RA) maps to the 2q33-q36 regi
A;Reference number: I37449; MUID:93252387; PMID:8486366

A;Accession: I37449
A;Molecule type: DNA
A;Residues: 1-350 <RES>
A;Cross-references: EMBL:X65858; NID:g312046; PIDN:CAA46688.1; PID:g312047
R;Ahuja, S.K.; Shetty, A.; Tiffany, H.L.; Murphy, P.M.
J. Biol. Chem. 269, 26381-26389, 1994
A;Title: Comparison of the genomic organization and promoter function for human interleu
A;Reference number: I37898; MUID:95014476; PMID:7929358
A;Accession: I38710
A;Molecule type: DNA
A;Residues: 1-350 <RE2>
A;Cross-references: EMBL:U11870; NID:g511804; PIDN:AAA64378.1; PID:g511805
A;Accession: I38711
A;Molecule type: mRNA
A;Residues: 1-16 <RE3>
A;Cross-references: EMBL:U11871; NID:g511806; PIDN:AAA64379.1; PID:g733002
R;Holmes, W.E.; Lee, J.; Kuang, W.J.; Rice, G.C.; Wood, W.I.
Science 253, 1278-1280, 1991
A;Title: Structure and functional expression of a human interleukin-8 receptor.
A;Reference number: A39445; MUID:91368199; PMID:1840701
A;Accession: A39445
A;Molecule type: mRNA
A;Residues: 1-275,'T',277-350 <HOL>
A;Cross-references: GB:M68932; NID:g186369; PIDN:AAA59159.1; PID:g186370
C;Genetics:
A;Gene: GDB:IL8RA
A;Cross-references: GDB:135039; OMIM:146929
A;Map position: 2q35-2q35
C;Superfamily: vertebrate rhodopsin
C;Keywords: G protein-coupled receptor; glycoprotein; membrane protein

Query Match   12.3%;   Score 194.5; DB 2;   Length 350;
Best Local Similarity   28.3%;   Pred. No. 7e-09;
Matches 82; Conservative 39; Mismatches 126; Indels 43; Gaps 11;

```
Qy   13 VAAFALGFPLNVLA--IRGATAHARLRITPSLVYALNLGCSDLLLTVSLPLKAVEALAS 69

Db   43 IIAYALVFLLSLLGNSLVMLVILYSRVGRSVTDVYLLNLALADLLFALTLPIWAASKV-- 100

Qy   70 GAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRPCYSWGVCAA 129

Db  101 NGWIFGTFLCKVVSLLKEVNFYSGILLLACISVDRYLAIVHATRTLTQKRHLVKF-VCLG 159

Qy  130 IWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAGPARFSLSLL 189

Db  160 CWGLSM-NLSLPFFLFRQA----------YHPNNSSPVCYEVLGNDTA-KWRMVLRIL 205

Qy  190 L----FFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPYNASNV 245

Db  206 PHTFGFIVPLFVMLFCYGFTLRTLFKAHM-GQKHRAMRVIFAVVLIFLLCWLPYNLVLL 263

Qy  246 ASFLY-----PNLGGSWRKLGL--ITGAWSVVLNPLVTGYLGR 281

Db  264 ADTLMRTQVIQESCERRRNIG---RALDATEILGFLHSCLNPIIYAFIGQ 310
```

Search completed: April 29, 2004, 12:12:04
Job time : 22 secs

```
                GenCore version 5.1.6
        Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:      April 29, 2004, 12:11:34 ; Search time 47 Seconds
                              (without alignments)
                              1769.295 Million cell updates/sec

Title:       US-10-202-687-2
Perfect score: 1584
Sequence:    1 MDLPPQLSFGLYVAAFALGF.........RGPGLKTVCAARTQGGKSQK 300

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5
Searched:    1138120 seqs, 277189581 residues

Total number of hits satisfying chosen parameters:      1138120

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     Published Applications AA:*
          1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
          2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
          3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
          4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
          5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
          6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
          7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
          8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
          9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
         10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
         11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
         12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
         13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
         14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
         15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
         16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
         17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
         18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

```
Result        Query
No.  Score  Match Length DB ID                        Description
---- -----  ----- ------ -- --                        -----------
  1  1584   100.0   300  14 US-10-251-385-250          Sequence 250, App
  2  1584   100.0   300  14 US-10-225-567A-475         Sequence 475, App
  3  1584   100.0   300  14 US-10-029-386-33080        Sequence 33080, A
  4  1584   100.0   300  15 US-10-202-687-2            Sequence 272, Appli
  5  1579   99.7    300  14 US-10-251-385-272          Sequence 272, App
  6   366   23.1    346  14 US-10-203-539-4            Sequence 4, Appli
  7   347   21.9    346  14 US-10-251-385-254          Sequence 254, App
  8   347   21.9    346  14 US-10-225-567A-605         Sequence 605, App
  9   347   21.9    346  14 US-10-203-539-2            Sequence 2, Appli
 10   347   21.9    401  14 US-10-029-386-33898        Sequence 33898, A
 11   342   21.6    346  14 US-10-251-385-274          Sequence 274, App
 12 340.5   21.5    330  14 US-10-348-190-2            Sequence 2, Appli
 13   334   21.1    330  14 US-10-348-190-4            Sequence 4, Appli
 14   332   21.0    330  14 US-10-251-385-258          Sequence 258, App
 15   332   21.0    330  14 US-10-225-567A-467         Sequence 467, App
 16   332   21.0    330  14 US-10-337-992-2            Sequence 2, Appli
 17   332   21.0    330  14 US-10-029-386-34068        Sequence 34068, A
 18   331   20.9    331  14 US-10-251-385-276          Sequence 276, App
 19   322   20.3    330   9 US-09-853-161-104          Sequence 104, App
 20   322   20.3    330   9 US-09-852-659A-104         Sequence 104, App
 21   322   20.3    330   9 US-09-852-797-104          Sequence 104, App
 22   322   20.3    330  12 US-10-058-993-104          Sequence 104, App
 23 298.5   18.8    319  14 US-10-203-539-6            Sequence 6, Appli
 24   244   15.4    420  12 US-10-081-810-41           Sequence 41, Appl
 25 231.5   14.6    385  12 US-10-081-810-43           Sequence 43, Appl
 26 231.5   14.6    385  14 US-10-225-567A-516         Sequence 516, App
 27 231.5   14.6    385  14 US-10-187-049-2            Sequence 2, Appli
 28   228   14.4    425   9 US-09-782-980-80           Sequence 80, Appl
 29   228   14.4    425   9 US-09-884-430-4            Sequence 4, Appli
 30   228   14.4    425  12 US-10-423-543-91           Sequence 91, Appli
 31   228   14.4    425  12 US-10-127-691-7            Sequence 7, Appli
 32   228   14.4    425  12 US-10-081-810-42           Sequence 42, App
 33   228   14.4    425  12 US-10-225-567A-324         Sequence 324, App
 34   228   14.4    425  14 US-10-336-489-4            Sequence 4, Appli
 35   228   14.4    425  14 US-10-177-293-130          Sequence 130, App
 36   228   14.4    425  14 US-10-176-464A-3           Sequence 3, Appli
 37   226   14.3    488  14 US-10-431-234-5            Sequence 5, Appli
 38   226   14.3    488  16 US-10-431-234-11           Sequence 11, Appl
 39 225.5   14.2    358  14 US-10-251-385-186          Sequence 186, App
 40 223.5   14.1    359  12 US-10-400-991-1            Sequence 1, Appli
 41 223.5   14.1    359  14 US-10-190-469-1            Sequence 1, Appli
 42 222.5   14.0    372  15 US-10-417-820A-149         Sequence 149, App
 43 221.5   14.0    359   9 US-09-739-151-2            Sequence 2, Appli
 44 221.5   14.0    359  10 US-09-782-974C-76          Sequence 76, Appli
 45 221.5   14.0    359  12 US-10-081-810-2            Sequence 2, Appli
```

ALIGNMENTS

```
RESULT 1
US-10-251-385-250
; Sequence 250, Application US/10251385
; Publication No. US2003010529241
; GENERAL INFORMATION:
;  APPLICANT: Behan, Dominic P.
;  APPLICANT: Chalmers, Derek T.
;  APPLICANT: Liaw, Chen W.
;  TITLE OF INVENTION: No. US2003010529241-Endogenous, Constitutively Activated Human G
;  TITLE OF INVENTION: Protein-Coupled
;  TITLE OF INVENTION: Receptors
;  FILE REFERENCE: AREN-0040
;  CURRENT APPLICATION NUMBER: US/10/251,385
;  CURRENT FILING DATE: 2002-09-20
;  PRIOR APPLICATION NUMBER: US/09/170,496
;  PRIOR FILING DATE: 1998-10-13
;  NUMBER OF SEQ ID NOS: 294
;  SOFTWARE: Patentin version 3.1
;  SEQ ID NO 250
;  LENGTH: 300
;  TYPE: PRT
;  ORGANISM: Homo sapiens
US-10-251-385-250

  Query Match             100.0%;  Score 1584;  DB 14;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 4.4e-139;
  Matches  300;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      1 MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRTPSLVYALNLGCSDLLLTVSLP 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRTPSLVYALNLGCSDLLLTVSLP 60

Qy     61 LKAVEALASGAWPLPASLCPVFAVAHFPPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRP 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 LKAVEALASGAWPLPASLCPVFAVAHFPPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRP 120

Qy    121 CYSWGVCAATWALVLCHLGLVFGLEAPGGWLDKSNTSLGINTPVNGSPVCLEAWDPASAG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   121 CYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG 180

Qy   181 PARFSLSLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 PARFSLSLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY 240

Qy   241 NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK 300


RESULT 2
US-10-225-567A-475
; Sequence 475, Application US/10225567A
; Publication No. US2003011379BA1
GENERAL INFORMATION:
;    APPLICANT: LifeSpan Biosciences
;    APPLICANT: Brown, Joseph P.
;    APPLICANT: Burmer, Glenna C.
;    APPLICANT: Roush, Christine L.
;    TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS
;    FILE REFERENCE: 1920-4-4
;    CURRENT APPLICATION NUMBER: US/10/225,567A
;    CURRENT FILING DATE: 2001-12-19
;    PRIOR APPLICATION NUMBER: 60/257,144
;    PRIOR FILING DATE: 2000-12-19
;    NUMBER OF SEQ ID NOS: 2292
;    SOFTWARE: PatentIn version 3.1
;
; SEQ ID NO 475
;    LENGTH: 300
;    TYPE: PRT
;    ORGANISM: Homo sapiens
US-10-225-567A-475

  Query Match             100.0%;  Score 1584;  DB 14;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 4.4e-139;
  Matches  300;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     1 MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLP 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLP 60

Qy    61 LKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRP 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 LKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRP 120

Qy   121 CYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG 180

Qy   181 PARFSLSLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 PARFSLSLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY 240

Qy   241 NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK 300


RESULT 3
US-10-029-386-33080
; Sequence 33080, Application US/10029386
; Publication No. US2003019470AA1
GENERAL INFORMATION:
;    APPLICANT: Penn, Sharron G.
;    APPLICANT: Rank, David R.
;    APPLICANT: Hanzel, David K.
;    TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
;    TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
;    FILE REFERENCE: AEOMICA-X-2
;    CURRENT APPLICATION NUMBER: US/10/029,386
;    CURRENT FILING DATE: 2001-12-20
;    NUMBER OF SEQ ID NOS: 34288
```

```
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;
; SEQ ID NO 33080
;    LENGTH: 300
;    TYPE: PRT
;    ORGANISM: Homo sapiens
;    FEATURE:
;    OTHER INFORMATION: MAP TO U62631.1
;    OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
;    OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.89
;    OTHER INFORMATION: SWISSPROT HIT: O14842, EVALUE 0.00e+00
US-10-029-386-33080

  Query Match             100.0%;  Score 1584;  DB 14;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 4.4e-139;
  Matches  300;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     1 MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLP 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLP 60

Qy    61 LKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRP 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 LKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRP 120

Qy   121 CYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG 180

Qy   181 PARFSLSLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 PARFSLSLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY 240

Qy   241 NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK 300


RESULT 4
US-10-202-687-2
; Sequence 2, Application US/10202687
; Publication No. US2004001910BA1
GENERAL INFORMATION:
;    APPLICANT: OWMAN, CHRISTER
;    APPLICANT: OLDE, BJORN
;    APPLICANT: KOTARSKY, KNUT
;    APPLICANT: NILSSON, NICLAS
;    APPLICANT: FLODGREN, ERIK
;    TITLE OF INVENTION: METHODS OF IDENTIFYING COMPOUNDS AFFECTING FATTY ACID
;    TITLE OF INVENTION: METABOLISM
;    FILE REFERENCE: 07675.0007 SEQUENCE LISTING
;    CURRENT APPLICATION NUMBER: US/10/202,687
;    CURRENT FILING DATE: 2002-07-24
;    NUMBER OF SEQ ID NOS: 4
;    SOFTWARE: PatentIn Ver. 2.1
;
; SEQ ID NO 2
;    LENGTH: 300
;    TYPE: PRT
;    ORGANISM: Homo sapiens
US-10-202-687-2

  Query Match             100.0%;  Score 1584;  DB 15;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 4.4e-139;
  Matches  300;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     1 MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLP 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLP 60

Qy    61 LKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRP 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 LKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRP 120

Qy   121 CYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG 180
```

Left column

```
Db  121 |||||||||||||||  CYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY 180
Qy  181                  PARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY 240
Db  181 ||||||||||||||  PARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY 240
Qy  241                  NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK 300
Db  241 ||||||||||||||  NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK 300
```

RESULT 5
US-10-251-385-272
; Sequence 272, Application US/10251385
; Publication No. US20030105292A1
; GENERAL INFORMATION:
; APPLICANT: Behan, Dominic P.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: No. US20030105292A1-Endogenous, Constitutively Activated Human G
; TITLE OF INVENTION: Protein-Coupled
; TITLE OF INVENTION: Receptors
; FILE REFERENCE: AREN-0040
; CURRENT APPLICATION NUMBER: US/10/251,385
; CURRENT FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: US/09/170,496
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 272
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-251-385-272

Query Match   99.7%; Score 1579; DB 14; Length 300;
Best Local Similarity 99.7%; Pred. No. 1.3e-138;
Matches 299; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy  1   MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLP 60
Db  1   MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLP 60
Qy  61  LKAVEALASGAWPLPASLCPVFAVAHFPPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRP 120
Db  61  LKAVEALASGAWPLPASLCPVFAVAHFPPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRP 120
Qy  121 CYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY 180
Db  121 CYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY 180
Qy  181 PARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY 240
Db  181 PARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY 240
Qy  241 NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK 300
Db  241 NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK 300
```

RESULT 6
US-10-203-539-4
; Sequence 4, Application US/10203539
; Publication No. US20030113810A1
; GENERAL INFORMATION:
; APPLICANT: GLAXO GROUP LIMITED
; TITLE OF INVENTION: NOVEL ASSAY
; FILE REFERENCE: PG3849USw
; CURRENT APPLICATION NUMBER: US/10/203,539
; CURRENT FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: GB 0003900.8
; PRIOR FILING DATE: 2000-02-18

Right column

; PRIOR APPLICATION NUMBER: GB 0007015.1
; PRIOR FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 4
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-203-539-4

Query Match   23.1%; Score 366; DB 14; Length 346;
Best Local Similarity 34.1%; Pred. No. 1.1e-25;
Matches 101; Conservative 38; Mismatches 129; Indels 28; Gaps 8;

```
Qy  1   MDLPPQLS--------FGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCS 51
Db  1   MDTGPDQSYFSGNHWFVFSVYLLTFLVGLPLNLLALVVFVGKLRCRPVAVDVLLNLTAS 60
Qy  52  DLLLTVSLPLKAVERALASGAWPLPASLCPVFAVAHFPPLYAGGGFLAALSAGRYLGAAFP 111
Db  61  DLLLLFLPFRMVEAANGMHWLPFILCPLSGFIFTTIYLTALFLAAVSIERFLSVAHP 120
Qy  112 LGYQAFRRPCYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVC- 170
Db  121 LWYKTRPRLGQAGLVSVACWLLASAHCSVVVIEFSGD-ISHSQGTNG--------TCY 170
Qy  171 LEAWDP--ASAGPARPSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGG 228
Db  171 LEFWKDQLAILLPVRLEMAVVLFVVPLIITSYCYSRLVWIIGRGG-SHRRQRRVAGLVAA 229
Qy  229 ALLTLLLCVGPYNASNVASFLYPNLGGS--WRKLGLITGAWSVVLNPLVTGYLGRG 282
Db  230 TLLNPLVCFGPYNVSHVVGYI---CGESPVWRIYVTLLSTLNSCVDPFVYFSSSG 282
```

RESULT 7
US-10-251-385-254
; Sequence 254, Application US/10251385
; Publication No. US20030105292A1
; GENERAL INFORMATION:
; APPLICANT: Behan, Dominic P.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: No. US20030105292A1-Endogenous, Constitutively Activated Human G
; TITLE OF INVENTION: Protein-Coupled
; TITLE OF INVENTION: Receptors
; FILE REFERENCE: AREN-0040
; CURRENT APPLICATION NUMBER: US/10/251,385
; CURRENT FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: US/09/170,496
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 254
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-251-385-254

Query Match   21.9%; Score 347; DB 14; Length 346;
Best Local Similarity 32.3%; Pred. No. 6.6e-24;
Matches 95; Conservative 41; Mismatches 134; Indels 24; Gaps 6;

```
Qy  1   MDLPPQLS--------FGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCS 51
Db  1   MDTGPDQSYFSGNHWFVFSVYLLTFLVGLPLNLLALVVFVGKLQRRPVAVDVLLNLTAS 60
Qy  52  DLLLTVSLPLKAVERALASGAWPLPASLCPVFAVAHFPPLYAGGGFLAALSAGRYLGAAFP 111
Db  61  DLLLLFLPFRMVEAANGMHWPLPFILCPLSGFIFTTIYLTALFLAAVSIERFLSVAHP 120
Qy  112 LGYQAFRRPCYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCL 171
```

```
Db      121 LWYKTRPRLGQAGLVSVACWLLASAHCSVVYVIEFSGD-ISHSQGTNG---------TCY 170

Qy      172 EAW---DPASAGPARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGG 228
                ||||   ::|| :||| |  || :  ||
Db      171 LEFRKDQLAILLPVRLEMAVVLFVVPLIITSYCYSRLVWILGRGG-SHRRQRRVAGLLAA 229

Qy      229 ALLTLLLCVGPYNASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRG 282
                || || |||| :: ::  || :  :: || ::|||
Db      230 TLLNFLVCFGPYNVSHVVGYICGE-SPAWRIYVTLLSTLNSCVDPFVYFSSSG 282


RESULT 8
US-10-225-567A-605
; Sequence 605, Application US/10225567A
; Publication No.US20030113798A1
; GENERAL INFORMATION:
;  APPLICANT: LifeSpan Biosciences
;  APPLICANT: Brown, Joseph P.
;  APPLICANT: Burmer, Glenna C.
;  APPLICANT: Roush, Christine L.
;  TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 1920-4-4
;  CURRENT APPLICATION NUMBER: US/10/225,567A
;  CURRENT FILING DATE: 2001-12-19
;  PRIOR APPLICATION NUMBER: 60/257,144
;  PRIOR FILING DATE: 2000-12-19
;  NUMBER OF SEQ ID NOS: 2292
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 605
;  LENGTH: 346
;  TYPE: PRT
;  ORGANISM: Homo sapiens
US-10-225-567A-605

Query Match             21.9%;  Score 347;  DB 14;  Length 346;
Best Local Similarity   32.3%;  Pred. No. 6.6e-24;
Matches   95; Conservative  41; Mismatches 134;  Indels  24;  Gaps   6;

Qy        1 MDLPPQLS--------FGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCS 51
              || ||:    :||| :||:| |  :: | |    ::|||
Db        1 MDTGPDQSYFSGNHWFVFSVYLLTFLVGLPLNLLALVVFVGKLQRRPVAVDVLLNLTAS 60

Qy       52 DLLLTVSLPLKAVERALASGAWPLPASLCPVFAVAHFPPLYAGGGFLAALSAGRYLGAAFP 111
              ||||||  ::| |||  :|||
Db       61 DLLLLLFLPFRMVEAANGMHWPLPFILCPLSGFIFFTTIYLTALFLAAVSIERFLSVAHP 120

Qy      112 LGYQAFRRPCYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCL 171
              | :|   ::|| :|
Db      121 LWYKTRPRLGQAGLVSVACWLLASAHCSVVYVIEFSGD-ISHSQGTNG---------TCY 170

Qy      172 EAW---DPASAGPARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGG 228
                ||||   ::|| :||| |  || :  ||
Db      171 LEFRKDQLAILLPVRLEMAVVLFVVPLIITSYCYSRLVWILGRGG-SHRRQRRVAGLLAA 229

Qy      229 ALLTLLLCVGPYNASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRG 282
                || || |||| :: ::  || :  :: || ::|||
Db      230 TLLNFLVCFGPYNVSHVVGYICGE-SPAWRIYVTLLSTLNSCVDPFVYFSSSG 282


RESULT 9
US-10-203-539-2
; Sequence 2, Application US/10203539
; Publication No. US20030113810A1
; GENERAL INFORMATION:
;  APPLICANT: GLAXO GROUP LIMITED
;  TITLE OF INVENTION: NOVEL ASSAY
;  FILE REFERENCE: PG3849USw
;  CURRENT APPLICATION NUMBER: US/10/203,539
;  CURRENT FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: GB 0003900.8
;  PRIOR FILING DATE: 2000-02-18
;  PRIOR APPLICATION NUMBER: GB 0007015.1
;  PRIOR FILING DATE: 2000-03-22
```

```
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 2
;  LENGTH: 346
;  TYPE: PRT
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: TRANSMEM
;  LOCATION: (18)..(41)
;  FEATURE:
;  NAME/KEY: TRANSMEM
;  LOCATION: (52)..(73)
;  FEATURE:
;  NAME/KEY: TRANSMEM
;  LOCATION: (88)..(111)
;  FEATURE:
;  NAME/KEY: TRANSMEM
;  LOCATION: (132)..(153)
;  FEATURE:
;  NAME/KEY: TRANSMEM
;  LOCATION: (188)..(212)
;  FEATURE:
;  NAME/KEY: TRANSMEM
;  LOCATION: (229)..(250)
;  FEATURE:
;  NAME/KEY: TRANSMEM
;  LOCATION: (259)..(278)
; US-10-203-539-2

Query Match             21.9%;  Score 347;  DB 14;  Length 346;
Best Local Similarity   32.3%;  Pred. No. 6.6e-24;
Matches   95; Conservative  41; Mismatches 134;  Indels  24;  Gaps   6;

Qy        1 MDLPPQLS--------FGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCS 51
              || ||:    :||| :||:| |  :: | |    ::|||
Db        1 MDTGPDQSYFSGNHWFVFSVYLLTFLVGLPLNLLALVVFVGKLQRRPVAVDVLLNLTAS 60

Qy       52 DLLLTVSLPLKAVERALASGAWPLPASLCPVFAVAHFPPLYAGGGFLAALSAGRYLGAAFP 111
              ||||||  ::| |||  :|||
Db       61 DLLLLLFLPFRMVEAANGMHWPLPFILCPLSGFIFFTTIYLTALFLAAVSIERFLSVAHP 120

Qy      112 LGYQAFRRPCYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCL 171
              | :|   ::|| :|
Db      121 LWYKTRPRLGQAGLVSVACWLLASAHCSVVYVIEFSGD-ISHSQGTNG---------TCY 170

Qy      172 EAW---DPASAGPARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGG 228
                ||||   ::|| :||| |  || :  ||
Db      171 LEFRKDQLAILLPVRLEMAVVLFVVPLIITSYCYSRLVWILGRGG-SHRRQRRVAGLLAA 229

Qy      229 ALLTLLLCVGPYNASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRG 282
                || || |||| :: ::  || :  :: || ::|||
Db      230 TLLNFLVCFGPYNVSHVVGYICGE-SPAWRIYVTLLSTLNSCVDPFVYFSSSG 282


RESULT 10
US-10-029-386-33898
; Sequence 33898, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT: Rank, David R.
;  APPLICANT: Hanzel, David K.
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GE
;  TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
;  FILE REFERENCE: AEOMICA-X-2
;  CURRENT APPLICATION NUMBER: US/10/029,386
;  CURRENT FILING DATE: 2001-12-20
;  NUMBER OF SEQ ID NOS: 34288
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33898
;  LENGTH: 401
;  TYPE: PRT
;  ORGANISM: Homo sapiens
```

```
;   FEATURE:
;     OTHER INFORMATION: MAP TO U62631.1
;     OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.4
;     OTHER INFORMATION: SWISSPROT HIT: O14843, EVALUE 0.00e+00
US-10-029-386-33898

Query Match          21.9%;  Score 347;  DB 14;  Length 401;
Best Local Similarity 32.3%;  Pred. No. 7.9e-24;
Matches    95;  Conservative  41;  Mismatches  134;  Indels  24;  Gaps  6;

Qy      1  MDLPPQLS---------FGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCS   51
                  ||         : |:| :: ||:|:|
Db     56  MDTGPDQSYFSGNHWFVFSVYLLTFLVGLPLNLLALVVFVGKLQRRPVAVDVLLINLTAS  115

Qy     52  DLLLTVSLPLKAVERALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFP  111
                 ||:|:||| ||||
Db    116  DLLLLLFLPFRMVEAANGMHWPLPFILCPLSGFIPFTTIYLTALFLAAVSIERFLSVAHP  175

Qy    112  LGYQAFRRPCYSWGVCAATWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCL  171
                  |:||||||::  : |:.|::.| |  ||:.|::||
Db    176  LWYKTRPRLGQAGLVSVACWLLASAHCSVVYVIEFSGD-ISHSQGTNG---------TCY  225

Qy    172  EAW---DPASAGPARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGG  228
                :       :||     :||||  ::  |::|.|:| |: :|:|: ||::|:|  ||
Db    226  LEFRKDQLAILLPVRLEMAVVLFVVPLIITSYCYSRLVWILGRGG-SHRRQRRVAGLLAA  284

Qy    229  ALLTLLLCVGPYNASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRG        282
                | ||||||||||:::: :   :  || | | :.: |:   ||: :: |:
Db    285  TLLNFLVCFGPYNVSHVVGYICGE-SPAWRIYVTLLSTLNSCVDPFVYFSSSG        337
```

```
RESULT 11
US-10-251-385-274
; Sequence 274, Application US/10251385
; Publication No. US20030105292A1
; GENERAL INFORMATION:
;   APPLICANT: Behan, Dominic P.
;   APPLICANT: Chalmers, Derek T.
;   APPLICANT: Liaw, Chen W.
;   TITLE OF INVENTION: No. US2003010529A1-Endogenous, Constitutively Activated Human G
;   TITLE OF INVENTION: Protein-Coupled
;   TITLE OF INVENTION: Receptors
;   FILE REFERENCE: AREN-0040
;   CURRENT APPLICATION NUMBER: US/10/251,385
;   CURRENT FILING DATE: 2002-09-20
;   PRIOR APPLICATION NUMBER: US/09/170,496
;   PRIOR FILING DATE: 1998-10-13
;   NUMBER OF SEQ ID NOS: 294
;   SOFTWARE: PatentIn version 3.1
;   SEQ ID NO 274
;   LENGTH: 346
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-251-385-274

Query Match          21.6%;  Score 342;  DB 14;  Length 346;
Best Local Similarity 32.0%;  Pred. No. 1.9e-23;
Matches    94;  Conservative  41;  Mismatches  135;  Indels  24;  Gaps  6;

Qy      1  MDLPPQLS---------FGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCS   51
                  ||         : |:| :: ||:|:|
Db      1  MDTGPDQSYFSGNHWFVFSVYLLTFLVGLPLNLLALVVFVGKLQRRPVAVDVLLINLTAS   60

Qy     52  DLLLTVSLPLKAVERALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFP  111
                 ||:|:||| ||||
Db     61  DLLLLLFLPFRMVEAANGMHWPLPFILCPLSGFIPFTTIYLTALFLAAVSIERFLSVAHP  120

Qy    112  LGYQAFRRPCYSWGVCAATWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCL  171
                  |:||||||::  : |:.|::.| |  ||:.|::||
Db    121  LWYKTRPRLGQAGLVSVACWLLASAHCSVVYVIEFSGD-ISHSQGTNG---------TCY  170

Qy    172  EAW---DPASAGPARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGG  228
                :       :||     :||||  ::  |::|.|:|
Db    172  EAW---DPASAGPARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGG  228
```

```
Db    171  LEFRKDQLAILLPVRLEMAVVLFVVPLIITSYCYSRLVWILGRGG-SHRRQRRVKGLLAA  229

Qy    229  ALLTLLLCVGPYNASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRG        282
                | ||||||||||:::: :   :  || | | :.: |:   ||: :: |:
Db    230  TLLNFLVCFGPYNVSHVVGYICGE-SPAWRIYVTLLSTLNSCVDPFVYFSSSG        282
```

```
RESULT 12
US-10-348-190-2
; Sequence 2, Application US/10348190
; Publication No. US20030166150A1
; GENERAL INFORMATION:
;   APPLICANT: Renauld, Jean-Christophe
;   APPLICANT: Louahed, Jamila
;   APPLICANT: Levitt, Roy
;   APPLICANT: Nicolaides, Nicholas
;   APPLICANT: Dong, Qu
;   TITLE OF INVENTION: Asthma Associated Factors as Targets for Treating
;   TITLE OF INVENTION: Atopic Allergies Including Asthma and Related Disorders
;   FILE REFERENCE: 036870-5072
;   CURRENT APPLICATION NUMBER: US/10/348,190
;   CURRENT FILING DATE: 2003-01-22
;   PRIOR APPLICATION NUMBER: US/09/157,24
;   PRIOR FILING DATE: 1998-09-18
;   PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/059,510
;   PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-19
;   PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/032,224
;   PRIOR FILING DATE: EARLIER FILING DATE: 1996-12-02
;   NUMBER OF SEQ ID NOS: 27
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 330
;   TYPE: PRT
;   ORGANISM: Murinae gen. sp.
US-10-348-190-2

Query Match          21.5%;  Score 340.5;  DB 14;  Length 330;
Best Local Similarity 30.9%;  Pred. No. 2.5e-23;
Matches    99;  Conservative  45;  Mismatches  129;  Indels  47;  Gaps  11;

Qy     12  YVAAFALGFPLNVLAIRGATAHARL-RLTPSLVYALNLGCSDLLLTVSLPLKAVEALASG   70
                 | |||| |||: ||:| :: |:|||
Db     14  YILIFLTGLPANLLALRAFMGRVRQPOPAPVHILLLNLTLADLLLLLLPFRIVEAASNF   73

Qy     71  AWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRPCYSWGVCAAI  130
                 ||:| :: |:: :  |||| || :|| |: ||
Db     74  RWYLPKIVCALTGFGYSSIYCSTWLLAGISMERYLGVAFPVQYKLSRRPLY--GVIAAL  131

Qy    131  --WALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG---PARFS  185
                  |:|:.||: :|:||||:: :  |  ||::
Db    132  VAWIMSFGHCTIVIVQ-------YLNSTEQVGT-ENQITCYENFTQBQLDVVLPVRLE  182

Qy    186  LSLLLFFLPLAITAFCY---VGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPYNA  242
                 ||:|| |: |:  |:       :|| :|:::|:.|
Db    183  LCLVLFFVPMAVTIFCYWRFVWIMLTQPHVGAQRRR--RAVGLAVVTLLNFLVCFGPYNM  240

Qy    243  SNVASFLYPNLGGSWRKLGLITGAWSVVLNPLV--------------------TGYL     279
                 | :: || | :   :  |
Db    241  SHLVGF-YLRQSPSWRVEAVVFSSLNASLDPLLFYFSSSSVRRAFGKGLLIRNPASSML  299

Qy    280  GRGPGLKTVCAARTQGGKSQ     299
                 |||: :
Db    300  GRG-AKETVEGTKMDRGGSQ     318
```

```
RESULT 13
US-10-348-190-4
; Sequence 4, Application US/10348190
; Publication No. US20030166150A1
; GENERAL INFORMATION:
;   APPLICANT: Renauld, Jean-Christophe
;   APPLICANT: Louahed, Jamila
;   APPLICANT: Levitt, Roy
```

; APPLICANT: Nicolaides, Nicholas
; APPLICANT: Dong, Qu
; TITLE OF INVENTION: Asthma Associated Factors as Targets for Treating
; TITLE OF INVENTION: Atopic Allergies Including Asthma and Related Disorders
; FILE REFERENCE: 036870-5072
; CURRENT APPLICATION NUMBER: US/10/348,190
; CURRENT FILING DATE: 2003-01-22
; PRIOR APPLICATION NUMBER: US/09/157,24
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/059,510
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/032,224
; PRIOR FILING DATE: EARLIER FILING DATE: 1996-12-02
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-348-190-4

Query Match   21.1%; Score 334; DB 14; Length 330;
Best Local Similarity   31.0%; Pred. No. 1e-22;
Matches 93; Conservative 53; Mismatches 120; Indels 34; Gaps 12;

```
Qy   12 YVAAFALGFPLNVLAIRGATAHARL-RLTPSLVYALNLGCSDLLLTVSLPLKAVEALASG 70
        :| :|||| |::| |:| :: ||   || :::: |  | |||||:||| |||||||: ::
Db   14 YIIIFLTGLPANLLALRAFVGRIRQPQPAPVHILLLSLTLADLLLLLLPFKIIEAASNF 73

Qy   71 AWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRPCYSWGVCAAI 130
        ||||::|: |||  ||| ||  :|| |||||| |||||||: ||||||| ||:: ||:||
Db   74 RWYLPKVVCALTSFGFYSSIYCSTWLLAGISIERYLGVAFPVQYKLSRRPLY--GVIAAL 131

Qy  131 --WALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPV-CLEAWDPASAG---PARF 184
          |: | |:|| : :|| :: :        :|::||| ||  |   | ||:|:    |
Db  132 VAWVMSFGHCTIVIIVQ------YLNTTEQVR---SGNEITCYENFTDNQLDVVLPVRL 181

Qy  185 SLSLLFFLPLAITAFCY---VGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPYN 241
        .|:||:||:|:::|:||     | |  :  |:||||:|::   |  ||:|:|| ||||||
Db  182 ELCLVLFFIPMAVTIFCYWRFVWIMLSQPHVGAQRRR--RAVGLAVVTLLNFLVCFGPYN 239

Qy  242 ASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGR------GPGLKTVCAARTQG 295
        .||||:|| |:||||||||||:||||||| ::|:|  ||:       ||::| ||| |:|
Db  240 VSHLVGY-HQRKSPWWRSIAVVFSSLNASLDPLLFYFSSSVVRRAFGRGLQVL---RNQG 295
```

RESULT 14
US-10-251-385-258
; Sequence 258, Application US/10251385
; Publication No. US20030105292A1
GENERAL INFORMATION:
; APPLICANT: Behan, Dominic P.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: No. US20030105292A1-Endogenous, Constitutively Activated Human G
; TITLE OF INVENTION: Protein-Coupled
; TITLE OF INVENTION: Receptors
; FILE REFERENCE: AREN-0040
; CURRENT APPLICATION NUMBER: US/10/251,385
; CURRENT FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: US/09/170,496
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: PatentIn version 3.1
SEQ ID NO 258
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-251-385-258

Query Match   21.0%; Score 332; DB 14; Length 330;
Best Local Similarity   31.0%; Pred. No. 1.6e-22;
Matches 93; Conservative 53; Mismatches 120; Indels 34; Gaps 12;

```
Qy   12 YVAAFALGFPLNVLAIRGATAHARL-RLTPSLVYALNLGCSDLLLTVSLPLKAVEALASG 70
Db   14 YIIIFLTGLPANLLALRAFVGRIRQPQPAPVHILLLSLTLADLLLLLLPFKIIEAASNF 73
Qy   71 AWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRPCYSWGVCAAI 130
Db   74 RWYLPKVVCALTSFGFYSSIYCSTWLLAGISIERYLGVAFPVQYKLSRRPLY--GVIAAL 131
Qy  131 --WALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPV-CLEAWDPASAG---PARF 184
Db  132 VAWVMSFGHCTIVIIVQ------YLNTTEQVR---SGNEITCYENFTDNQLDVVLPVRL 181
Qy  185 SLSLLFFLPLAITAFCY---VGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPYN 241
Db  182 ELCLVLFFIPMAVTIFCYWRFVWIMLSQPHVGAQRRR--RAVGLAVVTLLNFLVCFGPYN 239
Qy  242 ASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGR------GPGLKTVCAARTQG 295
Db  240 VSHLVGY-HQRKSPWWRSIAVVFSSLNASLDPLLFYFSSSVVRRAFGRGLQVL---RNQG 295
```

RESULT 15
US-10-225-567A-467
; Sequence 467, Application US/10225567A
; Publication No. US2003011379 8A1
GENERAL INFORMATION:
; APPLICANT: Lifespan Biosciences
; APPLICANT: Brown, Joseph P.
; APPLICANT: Burmer, Glenna C.
; APPLICANT: Roush, Christine L.
; TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 1920-4-4
; CURRENT APPLICATION NUMBER: US/10/225,567A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/257,144
; PRIOR FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 2292
; SOFTWARE: PatentIn version 3.1
SEQ ID NO 467
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-225-567A-467

Query Match   21.0%; Score 332; DB 14; Length 330;
Best Local Similarity   31.0%; Pred. No. 1.6e-22;
Matches 93; Conservative 53; Mismatches 120; Indels 34; Gaps 12;

```
Qy   12 YVAAFALGFPLNVLAIRGATAHARL-RLTPSLVYALNLGCSDLLLTVSLPLKAVEALASG 70
Db   14 YIIIFLTGLPANLLALRAFVGRIRQPQPAPVHILLLSLTLADLLLLLLPFKIIEAASNF 73
Qy   71 AWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRPCYSWGVCAAI 130
Db   74 RWYLPKVVCALTSFGFYSSIYCSTWLLAGISIERYLGVAFPVQYKLSRRPLY--GVIAAL 131
Qy  131 --WALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPV-CLEAWDPASAG---PARF 184
Db  132 VAWVMSFGHCTIVIIVQ------YLNTTEQVR---SGNEITCYENFTDNQLDVVLPVRL 181
Qy  185 SLSLLFFLPLAITAFCY---VGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPYN 241
Db  182 ELCLVLFFIPMAVTIFCYWRFVWIMLSQPHVGAQRRR--RAVGLAVVTLLNFLVCFGPYN 239
Qy  242 ASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGR------GPGLKTVCAARTQG 295
Db  240 VSHLVGY-HQRKSPWWRSIAVVFSSLNASLDPLLFYFSSSVVRRAFGRGLQVL---RNQG 295
```

Search completed: April 29, 2004, 12:17:26
Job time : 48 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2004, 12:08:54 ; Search time 23 Seconds
(without alignments)
673.382 Million cell updates/sec

Title: US-10-202-687-2
Perfect score: 1584
Sequence: 1 MDLPPQLSFGLYVAAFALGF.........RGPGLKTVCAARTQGGKSQK 300

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query % Match	Length	DB	ID	Description
1	1584	100.0	300	4	US-09-170-496D-250	Sequence 250, App
2	1579	99.7	300	4	US-09-170-496D-272	Sequence 272, App
3	347	21.9	346	4	US-09-170-496D-254	Sequence 254, App
4	347	21.9	401	3	US-08-820-521-2	Sequence 2, Appli
5	347	21.9	401	3	US-09-217-101-3	Sequence 2, Appli
6	347	21.9	401	3	US-09-248-715-2	Sequence 2, Appli
7	342	21.6	346	4	US-09-170-496D-274	Sequence 274, App
8	340.5	21.5	330	3	US-09-187-710-2	Sequence 2, Appli
9	332	21.0	330	2	US-08-788-750-2	Sequence 2, Appli
10	332	21.0	330	4	US-09-170-496D-258	Sequence 258, App
11	331	20.9	330	4	US-09-170-496D-276	Sequence 276, App
12	322	20.3	330	4	US-09-152-060-104	Sequence 104, App
13	231.5	14.6	385	3	US-09-053-866-2	Sequence 2, Appli
14	231.5	14.6	385	4	US-09-479-130-2	Sequence 2, Appli
15	231.5	14.6	385	4	US-09-472-130A-2	Sequence 2, Appli
16	228	14.4	425	1	US-07-657-769B-69	Sequence 69, Appl
17	228	14.4	425	1	US-08-097-938-7	Sequence 7, Appl
18	228	14.4	425	1	US-08-313-553-13	Sequence 13, Appl
19	228	14.4	425	1	US-07-789-184-220	Sequence 220, App
20	228	14.4	425	1	US-08-476-000-7	Sequence 7, Appli
21	228	14.4	425	1	US-08-475-263-220	Sequence 220, App
22	228	14.4	425	1	US-08-472-840-7	Sequence 7, Appli
23	228	14.4	425	1	US-08-485-886-220	Sequence 220, App
24	228	14.4	425	2	US-08-477-362-220	Sequence 220, App
25	228	14.4	425	2	US-08-477-134-220	Sequence 220, App
26	228	14.4	425	2	US-08-911-320A-3	Sequence 3, Appli
27	228	14.4	425	2	US-08-476-976-7	Sequence 7, Appli
28	228	14.4	425	2	US-08-742-440A-7	Sequence 7, Appl
29	228	14.4	425	2	US-08-560-098A-57	Sequence 57, Appl
30	228	14.4	425	2	US-08-767-993-13	Sequence 13, Appl
31	228	14.4	425	3	US-08-473-489A-220	Sequence 220, App
32	228	14.4	425	3	US-08-474-410-7	Sequence 7, Appl
33	228	14.4	425	3	US-08-485-695-220	Sequence 220, App
34	228	14.4	425	3	US-09-217-101-3	Sequence 3, Appli
35	228	14.4	425	3	US-08-018-760-220	Sequence 220, App
36	228	14.4	425	3	US-08-486-673B-7	Sequence 7, Appli
37	225.5	14.2	358	4	US-09-170-496D-186	Sequence 186, App
38	216.5	13.7	358	4	US-09-170-496D-40	Sequence 40, Appl
39	213	13.4	358	3	US-09-041-545-2	Sequence 2, Appli
40	213	13.4	358	3	US-09-327-925-2	Sequence 2, Appli
41	210	13.3	357	5	PCT-US95-07180-3	Sequence 3, Appli
42	205.5	13.0	408	3	US-08-742-440A-6	Sequence 6, Appli
43	202.5	12.8	328	3	US-08-513-974B-56	Sequence 56, Appl
44	202.5	12.8	328	3	US-08-513-974B-380	Sequence 380, App
45	202.5	12.8	328	4	US-09-461-436B-56	Sequence 56, Appl

ALIGNMENTS

RESULT 1
US-09-170-496D-250
; Sequence 250, Application US/09170496D
; Patent No. 6555339
; GENERAL INFORMATION:
; APPLICANT: Behan, Dominic P.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: No. 6555339-Endogenous, Constitutively Activated Human G Protein-C
; TITLE OF INVENTION: Receptors
; FILE REFERENCE: AREN-0040
; CURRENT APPLICATION NUMBER: US/09/170,496D
; CURRENT FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SEQ ID NO 250
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-170-496D-250

Query Match 100.0%; Score 1584; DB 4; Length 300;
Best Local Similarity 100.0%; Pred. No. 3.2e-126;
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLP 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLP 60

Qy   61 LKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGFLAALSAGRYLGAAFPLGYQAFRRP 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGFLAALSAGRYLGAAFPLGYQAFRRP 120

Qy 121 CYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 CYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG 180

Qy 181 PARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 PARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY 240

Qy 241 NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK 300

RESULT 2
US-09-170-496D-272
; Sequence 272, Application US/09170496D
; Patent No. 6555339

```
Db    61  DLLLLFLPPRMVEAANGMHWPLPPILCPLSGFIFFTTIYITALFLAAVSIERFLSVAHP 120
          | | :| |   || || |    ||  |  |: |::: | | ||:|| | | |:|::
Qy   112  LGYQAFRRPCYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCL 171
          |  |:     |      |::|  ||: |  |||||||||||||||||||||||||||:
Db   121  LWYKTRPRLGQAGLVSVACWLLASAHCSVVYVIEFSGD-ISHSQGTNG------TCY 170
          |   |: ||| ::: :|  | ||  |  | :: |  ||| :| :
Qy   172  EAW--DPASAGPARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGG 228
          |    : :|  :||:|||:||||||||||:||: |||||||||||||||||| ::
Db   171  LEFRKDQLAILLPVRLEMAVVLFVVPLIITSYCYSRLVWILGRGG-SHRRQRRVAGLLAA 229
          :   | |   || |  |: |:|| || ||:|  || |: |:   :| ::|: :|
Qy   229  ALLTLLCVGPYNASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRG 282
          : |  | :| ||      ||||  :   ||: :: ::
Db   230  TLLNFIVCFGPYNVSHVVGYICGE-SPAWRIYVTLLSTLNSCVDPFVYYFSSSG 282
```

```
Query Match          21.9%;   Score 347;  DB 2;  Length 401;
Best Local Similarity  32.3%;   Pred. No. 1.4e-21;
Matches  95;  Conservative 41;  Mismatches 134;  Indels 24;  Gaps  6;

Qy    1  MDLPPQLS--------FGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCS 51
         |||||||         |:|  | |   | :|   | |  || ||| |:  ::| | ||:
Db    1  MDTGPDQSYFSGNHWFVFSVLLTFLVGLLPLNLLALVVFVGKLQRRPVAVDVLLNLTAS 115

Qy   52  DLLLTVSLPLKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFP 111
         |||| | | | | | ||: |  |  || | |  || | | | | ||| | :| || |
```

```
Query Match          99.7%;   Score 1579;  DB 4;  Length 300;
Best Local Similarity  99.7%;   Pred. No. 8.5e-126;
Matches 299;  Conservative  0;  Mismatches  1;  Indels  0;  Gaps  0;

Qy    1  MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLP 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLP 60

Qy   61  LKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRP 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  LKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRP 120

Qy  121  CYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLRALARSGLT 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  CYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLRALARSGLT 180

Qy  181  PARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLCVGPY 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  PARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAKWVAGGALLTLLCVGPY 240

Qy  241  NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK 300
```

```
Query Match          21.9%;   Score 347;  DB 4;  Length 346;
Best Local Similarity  32.3%;   Pred. No. 1.1e-21;
Matches  95;  Conservative 41;  Mismatches 134;  Indels 24;  Gaps  6;

Qy    1  MDLPPQLS--------FGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCS 51
         |||||||         |:|  | |   | :|   | |  || ||| |:  ::| | ||:
Db    1  MDTGPDQSYFSGNHWFVFSVLLTFLVGLLPLNLLALVVFVGKLQRRPVAVDVLLNLTAS 60

Qy   52  DLLLTVSLPLKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFP 111
         |||| | | | | | ||: |  |  || | |  || | | | | ||| | :| || |
```

Db 116 DLLLLFLPFRMVEAANGMHWPLPFILCPLSGFIFFTTIYLTALFLAAVSIERFLSVAHP 175
Qy 112 LGYQAFRRPCYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCL 171
Db 176 LWYKTRPRLGQAGLVSVACWLLASAHCSVVYVIEFSGD-ISHSQGTNG--------TCY 225
Qy 172 EAW--DPASAGPARFSLSLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGG 228
Db 226 LEFRKDQLAILLPVRLEMAVVLFVVPLIITSYCYSRLVWILGRGG-SHRRQRRVAGLLAA 284
Qy 229 ALLTLLLCVGPYNASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRG 282
Db 285 TLLNPLVCFGPYNVSHVVGYICGE-SPAWRIYVTLLSTLNSCVDPFVYFSSSG 337

RESULT 6
US-09-248-715-2
; Sequence 2, Application US/09248715
; Patent No. 6277960
; GENERAL INFORMATION:
;   APPLICANT: BERGSMA, DERK
;   APPLICANT: SATHE, GANESH M.
;   APPLICANT: FUETTERER, WENDY
;   APPLICANT: MAO, JOYCE
;   TITLE OF INVENTION: CDNA CLONE HNFDY20 THAT ENCODES
;   TITLE OF INVENTION: A NOVEL HUMAN 7-TRANSMEMBRANE RECEPTOR
;   NUMBER OF SEQUENCES: 2
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Ratner & Prestia
;   STREET: P.O. Box 980
;   CITY: Valley Forge
;   STATE: PA
;   COUNTRY: USA
;   ZIP: 19482
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/248,715
;   FILING DATE: 09-Feb-1999
;   CLASSIFICATION: UNKNOWN
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/820,521
;   FILING DATE: 19-MAR-1997
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Prestia, Paul F
;   REGISTRATION NUMBER: 23,031
;   REFERENCE/DOCKET NUMBER: GP-50011-1
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 610-407-0700
;   TELEFAX: 610-407-0700
;   TELEX: 846169
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 401 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-248-715-2

Query Match   21.9%; Score 347; DB 3; Length 401;
Best Local Similarity 32.3%; Pred. No. 1.4e-21;
Matches 95; Conservative 41; Mismatches 134; Indels 24; Gaps 6;

Qy 1 MDLPPQLS--------FGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCS 51
Db 56 MDTGPDQSYFSGNHWFVFSVYLLTFLVGLPLNLLALVVFVGKLQRRPVAVDVLLNLTAS 115
Qy 52 DLLLTVSLPLKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFP 111

Db   116 DLLLLLFLPFRMVEAANGMHWPLPFILCPLSGFIFFTTIYLTALFLAAVSIERFLSVAHP 175

Qy   112 LGYQAFRRPCYSWGVCAAIWALVLCHLGLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCL 171

Db   176 LWYKTRPRLGQAGLVSVACWLLASAHCSVYYVIEFSGD-ISHSQGTNG--------TCY 225

Qy   172 EAW---DPASAGPARFSLSLLLFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGG 228

Db   226 LEFRKDQLAILLPVRLEMAVVLFVVPLIITSYCYSRLVWILGRCG-SHRRQRRVAGLLAA 284

Qy   229 ALLTLLLCVGPYNASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRG 282

Db   285 TLLNFLVCFGPYNVSHVVGYICGE-SPAWRIYVTLLSTLNSCVDPFVYFSSSSG 337

RESULT 7
US-09-170-496D-274
; Sequence 274, Application US/09170496D
; Patent No. 6555339
; GENERAL INFORMATION:
; APPLICANT: Behan, Dominic P.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: No. 6555339-Endogenous, Constitutively Activated Human G Protein-
; TITLE OF INVENTION: Receptors
; FILE REFERENCE: AREN--0040
; CURRENT APPLICATION NUMBER: US/09/170,496D
; CURRENT FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 274
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-170-496D-274

Query Match   21.6%; Score 342; DB 4; Length 346;
Best Local Similarity   32.0%; Pred. No. 3e-21;
Matches 94; Conservative 41; Mismatches 135; Indels 24; Gaps   6;

Qy   1 --------FGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCS 51

Db   1 MDLPPQLS---------FGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCS 51

Db   1 MDTGPDQSYFSGNHWFVFSVYLLTFLVGLPLNLLALVFVGKLQRRPVAVDVLLNLTAS 60

Qy   52 DLLLTVSLPLKAVEALASGAWPLPASLCPVFAVAHFPLYAGGGFLAALSAGRYLGAAFP 111

Db   61 DLLLLLFLPFRMVEAANGMHWPLPFILCPLSGFIFFTTIYLTALFLAAVSIERFLSVAHP 120

Qy   112 LGYQAFRRPCYSWGVCAAIWALVLCHLGLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCL 171

Db   121 LWYKTRPRLGQAGLVSVACWLLASAHCSVYYVIEFSGD-ISHSQGTNG--------TCY 170

Qy   172 EAW---DPASAGPARFSLSLLLFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGG 228

Db   171 LEFRKDQLAILLPVRLEMAVVLFVVPLIITSYCYSRLVWILGRGG-SHRRQRRVKGLLAA 229

Qy   229 ALLTLLLCVGPYNASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRG 282

Db   230 TLLNFLVCFGPYNVSHVVGYICGE-SPAWRIYVTLLSTLNSCVDPFVYFSSSSG 282

RESULT 8
US-09-187-710-2
; Sequence 2, Application US/09187710A
; Patent No. 6180365
; GENERAL INFORMATION:
; APPLICANT: LANE, PAMELA
; APPLICANT: TSUI, PING
; APPLICANT: ELSHOURBAGY, NABIL
; APPLICANT: VAWTER, LISA
; TITLE OF INVENTION: MOUSE 7-TRANSMEMBRANE RECEPTOR GPR43
; FILE REFERENCE: GP-70566

; CURRENT APPLICATION NUMBER: US/09/187,710A
; CURRENT FILING DATE: 1998-11-06
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 330
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-09-187-710-2

Query Match   21.5%; Score 340.5; DB 3; Length 330;
Best Local Similarity   30.9%; Pred. No. 3.8e-21;
Matches 99; Conservative 45; Mismatches 129; Indels 47; Gaps 11;

Qy   12 YVAARAFALGFPLNVLAIRGATAHARL-RLTPSLVYALNLGCSDLLLTVSLPLKAVEALASG 70

Db   14 YILIFLTGLPANLLALRAFMGRVRQPQAPVHILLLNLTLADLLLLLLPFRIVEAASNF 73

Qy   71 AWPLPASLCPVFAVAHFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRPCYSWGVCAAI 130

Db   74 RWYLPKIVCALTGFGFYSSIYCSTWLLAGISMERYLGVAPPVQYKLSRRPLY--GVIAAL 131

Qy   131 --WALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG---PARFS 185

Db   132 VAWTMSFGHCTIVIVQ------YLNSTEQVGT--ENQITCYENFTQEQLDVVLPVRLE 182

Qy   186 LSLLLFFLPLAITAPCY---VGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPYNA 242

Db   183 LCLVLFFVPMAVTIFCYWRFVWIMLTQPHVGAQRRR--RAVGLAVVTLLNFLVCFGPYNM 240

Qy   243 SNVASFLYPNLGGSWRKLGLITGAWSVVLNPLV--------------------TGYL 279

Db   241 SHLVGF-YLRQSPSRVEAVVFSSLNASLDPLLFYFSSSSVVRRAFGKGLLLIRNPASSVL 299

Qy   280 GRGPGLKTVCAARTQGGKSQ 299

Db   300 GRG-AKETVEGTKMDRGGSQ 318

RESULT 9
US-08-788-750-2
; Sequence 2, Application US/08788750
; Patent No. 5910430
; GENERAL INFORMATION:
; APPLICANT: Ellis, Catherine
; APPLICANT: Bergsma, Derk
; TITLE OF INVENTION: No. 5910430el G-Protein Coupled Receptor
; TITLE OF INVENTION: (HTADX50)
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/788,750
; FILING DATE: 24-JAN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Han, William T
; REGISTRATION NUMBER: 34,344
; REFERENCE/DOCKET NUMBER: ATG50048
; TELECOMMUNICATION INFORMATION:

```
;    TELEPHONE: 610-270-5219
;    TELEFAX: 610-270-4026
;    TELEX:
INFORMATION FOR SEQ ID NO: 2:
;    SEQUENCE CHARACTERISTICS:
;    LENGTH: 330 amino acids
;    TYPE: amino acid
;    STRANDEDNESS: single
;    TOPOLOGY: linear
US-08-788-750-2

Query Match             21.0%;  Score 332; DB 2; Length 330;
Best Local Similarity   31.0%;  Pred. No. 2e-20;
Matches  93; Conservative  53; Mismatches 120; Indels 34; Gaps 12;

Qy   12 YVAAFALGFPLNVLAIRGATAHARL-RLTPSLVYALNLGCSDLLLTVSLPLKAVEALASG 70
Db   14 YIIIFLTGLPANLLALRAFVGRIRQPQPAPVHILLSLTLADLLLLLLLPFKIIEAASNF 73
Qy   71 AWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRPCYSWGVCAAI 130
Db   74 RWYLPKVVCALTSFGFYSSIYCSTWLLAGISIERYLGVAFPVQYKLSRRPLY--GVIAAL 131
Qy  131 --WALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPV-CLEAWDPASAG---PARF 184
Db  132 VAWVMSFGHCTIVIIVQ------YLNTTEQVR---SGNEITCYENFTDNQLDVVLPVRL 181
Qy  185 SLSLLFFLPLAITAFCY---VGCLRALARSGLTHRRKLRAAWVAGGALLTLLCVGPYN 241
Db  182 ELCLVLFFIPMAVTIFCYWRFVWIMLSQPLVGAQRRR--RAVGLAVVTLLNFLVCFGPYN 239
Qy  242 ASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGR------GPGLKTVCAARTQG 295
Db  240 VSHLVGY-HQRKSPWWRSIAVVFSSLNASLDPLLFYFSSSVVRRAFGRGLQVL---RNQG 295

RESULT 10
US-09-170-496D-258
; Sequence 258, Application US/09170496D
; Patent No. 6555339
; GENERAL INFORMATION:
; APPLICANT: Behan, Dominic P.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: No. 6555339-Endogenous, Constitutively Activated Human G Protein
; TITLE OF INVENTION: Receptors
; FILE REFERENCE: AREN-0040
; CURRENT APPLICATION NUMBER: US/09/170,496D
; CURRENT FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 258
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-170-496D-258

Query Match             21.0%;  Score 332; DB 4; Length 330;
Best Local Similarity   31.0%;  Pred. No. 2e-20;
Matches  93; Conservative  53; Mismatches 120; Indels 34; Gaps 12;

Qy   12 YVAAFALGFPLNVLAIRGATAHARL-RLTPSLVYALNLGCSDLLLTVSLPLKAVEALASG 70
Db   14 YIIIFLTGLPANLLALRAFVGRIRQPQPAPVHILLSLTLADLLLLLLLPFKIIEAASNF 73
Qy   71 AWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRPCYSWGVCAAI 130
Db   74 RWYLPKVVCALTSFGFYSSIYCSTWLLAGISIERYLGVAFPVQYKLSRRPLY--GVIAAL 131
Qy  131 --WALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPV-CLEAWDPASAG---PARF 184
Db  132 VAWVMSFGHCTIVIIVQ------YLNTTEQVR---SGNEITCYENFTDNQLDVVLPVRL 181
```

```
Qy  185 SLSLLFFLPLAITAFCY---VGCLRALARSGLTHRRKLRAAWVAGGALLTLLCVGPYN 241
Db  182 ELCLVLFFIPMAVTIFCYWRFVWIMLSQPLVGAQRRR--RAVGLAVVTLLNFLVCFGPYN 239
Qy  242 ASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGR------GPGLKTVCAARTQG 295
Db  240 VSHLVGY-HQRKSPWWRSIAVVFSSLNASLDPLLFYFSSSVVRRAFGRGLQVL---RNQG 295

RESULT 11
US-09-170-496D-276
; Sequence 276, Application US/09170496D
; Patent No. 6555339
; GENERAL INFORMATION:
; APPLICANT: Behan, Dominic P.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: No. 6555339-Endogenous, Constitutively Activated Human G Protein-(
; TITLE OF INVENTION: Receptors
; FILE REFERENCE: AREN-0040
; CURRENT APPLICATION NUMBER: US/09/170,496D
; CURRENT FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 276
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-170-496D-276

Query Match             20.9%;  Score 331; DB 4; Length 330;
Best Local Similarity   31.0%;  Pred. No. 2.4e-20;
Matches  93; Conservative  53; Mismatches 120; Indels 34; Gaps 12;

Qy   12 YVAAFALGFPLNVLAIRGATAHARL-RLTPSLVYALNLGCSDLLLTVSLPLKAVEALASG 70
Db   14 YIIIFLTGLPANLLALRAFVGRIRQPQPAPVHILLSLTLADLLLLLLLPFKIIEAASNF 73
Qy   71 AWPLPASLCPVAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRPCYSWGVCAAI 130
Db   74 RWYLPKVVCALTSFGFYSSIYCSTWLLAGISIERYLGVAFPVQYKLSRRPLY--GVIAAL 131
Qy  131 --WALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPV-CLEAWDPASAG---PARF 184
Db  132 VAWVMSFGHCTIVIIVQ------YLNTTEQVR---SGNEITCYENFTDNQLDVVLPVRL 181
Qy  185 SLSLLFFLPLAITAFCY---VGCLRALARSGLTHRRKLRAAWVAGGALLTLLCVGPYN 241
Db  182 ELCLVLFFIPMAVTIFCYWRFVWIMLSQPLVGAQRRR--RAKGLAVVTLLNFLVCFGPYN 239
Qy  242 ASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGR------GPGLKTVCAARTQG 295
Db  240 VSHLVGY-HQRKSPWWRSIAVVFSSLNASLDPLLFYFSSSVVRRAFGRGLQVL---RNQG 295

RESULT 12
US-09-152-060-104
; Sequence 104, Application US/09152060
; Patent No. 6448230
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 28 Human Secreted Proteins
; FILE REFERENCE: PZ003P1.US
; CURRENT APPLICATION NUMBER: US/09/152,060
; CURRENT FILING DATE: 1998-09-11
; EARLIER APPLICATION NUMBER: PCT/US98/04858
; EARLIER FILING DATE: 1998-03-12
; EARLIER APPLICATION NUMBER: 60/040,762
; EARLIER FILING DATE: 1997-03-14
; EARLIER APPLICATION NUMBER: 60/040,710
; EARLIER FILING DATE: 1997-03-14
; EARLIER APPLICATION NUMBER: 60/050,934
; EARLIER FILING DATE: 1997-05-30
```

```
;  ;   EARLIER APPLICATION NUMBER: 60/048,100
;  ;   EARLIER FILING DATE: 1997-05-30
;  ;   EARLIER APPLICATION NUMBER: 60/048,357
;  ;   EARLIER FILING DATE: 1997-05-30
;  ;   EARLIER APPLICATION NUMBER: 60/048,189
;  ;   EARLIER FILING DATE: 1997-05-30
;  ;   EARLIER APPLICATION NUMBER: 60/057,765
;  ;   EARLIER FILING DATE: 1997-09-05
;  ;   EARLIER APPLICATION NUMBER: 60/048,970
;  ;   EARLIER FILING DATE: 1997-06-06
;  ;   EARLIER APPLICATION NUMBER: 60/068,368
;  ;   EARLIER FILING DATE: 1997-12-19
;  ;   NUMBER OF SEQ ID NOS: 118
;  ;   SOFTWARE: PatentIn Ver. 2.0
;  ;
;  ;   SEQ ID NO 104
;  ;   LENGTH: 330
;  ;   TYPE: PRT
;  ;   ORGANISM: Homo sapiens
;  ;   FEATURE:
;  ;   NAME/KEY: SITE
;  ;   LOCATION: (7)
;  ;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;  ;   FEATURE:
;  ;   NAME/KEY: SITE
;  ;   LOCATION: (147)
;  ;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;  ;   FEATURE:
;  ;   NAME/KEY: SITE
;  ;   LOCATION: (181)
;  ;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;  ;   FEATURE:
;  ;   NAME/KEY: SITE
;  ;   LOCATION: (190)
;  ;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;  ;   FEATURE:
;  ;   NAME/KEY: SITE
;  ;   LOCATION: (260)
;  ;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-152-060-104

Query Match           20.3%;   Score 322;   DB 4;   Length 330;
Best Local Similarity  31.0%;   Pred. No. 1.4e-19;
Matches  93;  Conservative  50;  Mismatches 123;  Indels  34;  Gaps  12;

Qy   12  YVAAFALGFPLNVLAIRGATAHARL-RLTPSLVYALNLGCSDLLLTVSLPLKAVEALASG  70
             :|:|  |  ||: |:|    :| ::|||| :||||
Db   14  YIIIFLTGLPANLLALRAFVGRIRQPQAPVHILLSLTLADLLLLLLPFKIIEAASNF  73

Qy   71  AWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRPCYSWGVCAAI  130
             || |:   |     || : :|||| :||||
Db   74  RWYLPKVVCALTSFGFYSSIYCSTWLLAGISIERYLGVAFPVQYKLSRRPLY--GVIRAL  131

Qy  131  -WALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPV-CLEAWDPASAG---PARF  184
             |:|||                ::|:   |||
Db  132  VAWVMSFGHCTIVIIXQ------YLNTTEQVR---SGNEITCYENFTDNQLDVVLPVRX  181

Qy  185  SLSLLLFPLPLAITAFCY---VGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPYN  241
             :|||               ||:|
Db  182  ELCLVLFFXPMAVTIFCYWRFWIMLSQPLVGAQRRR--RAVGLAVVTLLNFLVCFGPYN  239

Qy  242  ASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGR------GPGLKTVCAARTGG  295
             |:: :|               ::||::: :
Db  240  VSHLVGY-HQRKSPWWRSIAVXFSSLNASLDPLLFYFSSSVVRRAFGRGLQVL--RNQG  295

RESULT 13
US-09-053-866-2
;  Sequence 2, Application US/09053866
;  Patent No. 6111075
;  GENERAL INFORMATION:
;  ;   APPLICANT: Xu, Wenfeng
;  ;   APPLICANT: Presnell, Scott R.
;  ;   APPLICANT: Yee, David P.
```

```
;  ;   APPLICANT: Foster, Donald C.
;  ;   TITLE OF INVENTION: PROTEASE-ACTIVATED RECEPTOR
;  ;   TITLE OF INVENTION: PAR4 (ZCHEMR2)
;  ;   NUMBER OF SEQUENCES: 12
;  ;   CORRESPONDENCE ADDRESS:
;  ;   ADDRESSEE: ZymoGenetics, Inc.
;  ;   STREET: 1201 Eastlake Avenue East
;  ;   CITY: Seattle
;  ;   STATE: WA
;  ;   COUNTRY: USA
;  ;   ZIP: 98102
;  ;   COMPUTER READABLE FORM:
;  ;   MEDIUM TYPE: Diskette
;  ;   COMPUTER: IBM Compatible
;  ;   OPERATING SYSTEM: DOS
;  ;   SOFTWARE: FastSEQ for Windows Version 2.0
;  ;   CURRENT APPLICATION DATA:
;  ;   APPLICATION NUMBER: US/09/053,866
;  ;   FILING DATE:
;  ;   PRIOR APPLICATION DATA:
;  ;   APPLICATION NUMBER:
;  ;   FILING DATE:
;  ;   ATTORNEY/AGENT INFORMATION:
;  ;   NAME: Leith, Debra K
;  ;   REGISTRATION NUMBER: 32,619
;  ;   REFERENCE/DOCKET NUMBER: 98-10
;  ;   TELECOMMUNICATION INFORMATION:
;  ;   TELEPHONE: 206-442-6674
;  ;   TELEFAX: 206-442-6678
;  ;   TELEX:
;  ;   INFORMATION FOR SEQ ID NO: 2:
;  ;   SEQUENCE CHARACTERISTICS:
;  ;   LENGTH: 385 amino acids
;  ;   TYPE: amino acid
;  ;   STRANDEDNESS: single
;  ;   TOPOLOGY: linear
;  ;   MOLECULE TYPE: protein
;  ;   FRAGMENT TYPE: internal
US-09-053-866-2

Query Match            14.6%;   Score 231.5;   DB 3;   Length 385;
Best Local Similarity  24.9%;   Pred. No. 7.3e-12;
Matches  88;  Conservative  29;  Mismatches 123;  Indels 113;  Gaps  8;

Qy    3  LPPQLSFGLYVAAPALGFPLNVLAIRG-ATAHARLRLTPSLVYALNLGCSDLLLTVSLPL  61
             :|:| ||:                ||::  |||||  :||||
Db   75  VPTRLVPALYGLVLVVGLPANGLALWVLATQAPRL--PSTMLLMNLATADLLLALALPP  131

Qy   62  KAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRPC  121
             |||||:|                  : :|||| :||||
Db  132  RIAYHLRGQRWPFGEAACRLATAALYGHMYGSVLLLAAVSLDRYLALVHPLRRARALRGRR  191

Qy  122  YSWGVCAAIWAL-----------------------VLCHLGLVFGLEAPGGWLDHSNTSLG  159
             || :                                ||||
Db  192  LALGLCMAAWLMAAALALPLTLQRQTFRLARSDRVLCHDALPLDAQA------------  238

Qy  160  INTPVNGSPVCLEAWDPASAGPARFSLSLLLFPLPLAITAFCYVGCLRALARSGLTHRRK  219
                          :        |||     :||||
Db  239  -------------SHWQPAFT-----CLALLGCFLPLLAMLLCYGATLHTLAASGRRYGHA  281

Qy  220  LR------------------------AAWVAGGALLTLLLCVGPYNAS  243
             |:                         ||||
Db  282  LRLTAVVLASAVAFFVPSNLLLLHYSDPSPSAWGNLYGAVVPSLALSTLNSCVDP----  337

Qy  244  NVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGRGPGLKTVCAARTQGG  296
             || :|  |                 :||::|
Db  338  ----FIYYVSAEFR---------------DKVRAGLFQRSPGDTVASKASAEGG  373

RESULT 14
US-09-479-130-2
;  Sequence 2, Application US/09479130
```

```
; Patent No. 6436400
; GENERAL INFORMATION:
;  APPLICANT: Xu, Wenfeng
;  APPLICANT: Presnell, Scott R.
;  APPLICANT: Yee, David P.
;  APPLICANT: Foster, Donald C.
;  TITLE OF INVENTION: PROTEASE-ACTIVATED RECEPTOR
;  TITLE OF INVENTION: PAR4 (ZCHEMR2)
;  NUMBER OF SEQUENCES: 12
;  CORRESPONDENCE ADDRESS:
;   ADDRESSEE: ZymoGenetics, Inc.
;   STREET: 1201 Eastlake Avenue East
;   CITY: Seattle
;   STATE: WA
;   COUNTRY: USA
;   ZIP: 98102
;  COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ for Windows Version 2.0
;  CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/479,130
;   FILING DATE:
;   CLASSIFICATION:
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
;  ATTORNEY/AGENT INFORMATION:
;   NAME: Leith, Debra K
;   REGISTRATION NUMBER: 32,619
;   REFERENCE/DOCKET NUMBER: 98-10
;  TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 206-442-6674
;   TELEFAX: 206-442-6678
;   TELEX:
; INFORMATION FOR SEQ ID NO: 2:
;  SEQUENCE CHARACTERISTICS:
;   LENGTH: 385 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;  MOLECULE TYPE: protein
;  FRAGMENT TYPE: internal
US-09-479-130-2

Query Match             14.6%;  Score 231.5;  DB 4;  Length 385;
Best Local Similarity   24.9%;  Pred. No. 7.3e-12;
Matches   88;  Conservative   29;  Mismatches 123;  Indels 113;  Gaps   8;

Qy     3 LPPQLSFGLYVAAFALGFPLNVLAIRG-ATAHARLRLTPSLVYALNLGCSDLLLTVSLPL 61
           :|:|  |||| ||    :|:||| :||: :||||||||||:::||| :||||| ::||
Db    75 VPTRLVPALYGLVLVVGLPANGLALMVLATQAPRL---PSTMLLMNLATADLLLALALPP 131

Qy    62 KAVEALASGAWPLPASLCPVFAVAHFPPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRPC 121
           : :|  |  |  | :  | :|  |   :|||||||:| | |:||| || ||| | ||:||
Db   132 RIAYHLRGQRWPFGEAACRLATAALYGHMYGSVLLAAVSLDRYLALVHPLRARALRGRR 191

Qy   122 YSWGVCAAIWAL------------------------------VLCHLGLVFGLEAPGGWLDHSNTSLG 159
           |:|| ||:|:                                 ||||:||| ||:||:|:
Db   192 LALGLCMAAWLMAAALALPLTLQRQTFRLARSDRVLCHDALPLDAQA------------- 238

Qy   160 INTPVNGSPVCLEAWDPASAGPARFSLSLLLFFLPLAITAFCVVGCLRALARSGLTHRRK 219
            |  | :|  ||                                  || :|  || |||
Db   239 ----------SHWQPAFT----CLALLGCFLPLLAMLLCYGATLHTLAASGRRYGHA 281

Qy   220 LR------------------------------------AAWVAGGALLTLLLCVGPYNAS 243
           ||                                      ::| || ||||||| |:
Db   282 LRLTAVVLASAVAFPVPSNLLLLHYSDPSPSAWGNLYGAYVPSLALSTLNSCVDP---- 337

Qy   244 NVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGG 296
            :||  |      :|  ||: :|  | |||||:||||  ||:|||| | :|||
Db   338 ---FIYYVSAEFR------------DKVRAGLFQRSPGDTVASKASAEGG 373
```

```
RESULT 15
US-09-472-130A-2
; Sequence 2, Application US/09472130A
; Patent No. 6473765
; GENERAL INFORMATION:
;  APPLICANT: Xu, Wenfeng
;  APPLICANT: Presnell, Scott R.
;  APPLICANT: Yee, David P.
;  APPLICANT: Foster, Donald C.
;  TITLE OF INVENTION: PROTEASE-ACTIVATED RECEPTOR PAR4
;  TITLE OF INVENTION: (ZCHEMR2)
;  FILE REFERENCE: 98-10D2
;  CURRENT APPLICATION NUMBER: US/09/472,130A
;  CURRENT FILING DATE: 2000-01-07
;  PRIOR APPLICATION NUMBER: US 09/053,866
;  PRIOR FILING DATE: 1998-04-01
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;  LENGTH: 385
;  TYPE: PRT
;  ORGANISM: Homo sapiens
US-09-472-130A-2

Query Match             14.6%;  Score 231.5;  DB 4;  Length 385;
Best Local Similarity   24.9%;  Pred. No. 7.3e-12;
Matches   88;  Conservative   29;  Mismatches 123;  Indels 113;  Gaps   8;

Qy     3 LPPQLSFGLYVAAFALGFPLNVLAIRG-ATAHARLRLTPSLVYALNLGCSDLLLTVSLPL 61
           :|:|  |||| ||    :|:||| :||: :||||||||||:::||| :||||| ::||
Db    75 VPTRLVPALYGLVLVVGLPANGLALMVLATQAPRL---PSTMLLMNLATADLLLALALPP 131

Qy    62 KAVEALASGAWPLPASLCPVFAVAHFPPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRPC 121
           : :|  |  |  | :  | :|  |   :|||||||:| | |:||| || ||| | ||:||
Db   132 RIAYHLRGQRWPFGEAACRLATAALYGHMYGSVLLAAVSLDRYLALVHPLRARALRGRR 191

Qy   122 YSWGVCAAIWAL------------------------------VLCHLGLVFGLEAPGGWLDHSNTSLG 159
           |:|| ||:|:                                 ||||:||| ||:||:|:
Db   192 LALGLCMAAWLMAAALALPLTLQRQTFRLARSDRVLCHDALPLDAQA------------- 238

Qy   160 INTPVNGSPVCLEAWDPASAGPARFSLSLLLFFLPLAITAFCVVGCLRALARSGLTHRRK 219
            |  | :|  ||                                  || :|  || |||
Db   239 ----------SHWQPAFT----CLALLGCFLPLLAMLLCYGATLHTLAASGRRYGHA 281

Qy   220 LR------------------------------------AAWVAGGALLTLLLCVGPYNAS 243
           ||                                      ::| || ||||||| |:
Db   282 LRLTAVVLASAVAFPVPSNLLLLHYSDPSPSAWGNLYGAYVPSLALSTLNSCVDP---- 337

Qy   244 NVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGG 296
            :||  |      :|  ||: :|  | |||||:||||  ||:|||| | :|||
Db   338 ---FIYYVSAEFR------------DKVRAGLFQRSPGDTVASKASAEGG 373

Search completed: April 29, 2004, 12:12:39
Job time : 25 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6

Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2004, 12:07:24 ; Search time 45 Seconds
             (without alignments)
             2103.455 Million cell updates/sec

Title: US-10-202-687-2
Perfect score: 1584
Sequence: 1 MDLPPQLSFGLYVAAFALGF.........RGPGLKTVCAARTQGGKSQK 300

Scoring table: BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   SPTREMBL_25:*
          1: sp_archea:*
          2: sp_bacteria:*
          3: sp_fungi:*
          4: sp_human:*
          5: sp_invertebrate:*
          6: sp_mammal:*
          7: sp_mhc:*
          8: sp_organelle:*
          9: sp_phage:*
         10: sp_plant:*
         11: sp_rodent:*
         12: sp_virus:*
         13: sp_vertebrate:*
         14: sp_unclassified:*
         15: sp_rvirus:*
         16: sp_bacteriap:*
         17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1329	83.9	300	11	Q8K3T5	Q8k3t5 mus musculu
2	1320	83.3	300	11	Q8K3T4	Q8k3t4 rattus norv
3	340.5	21.5	330	11	Q8VCK6	Q8vck6 mus musculu
4	222.5	14.0	396	11	Q8BZ77	Q8bz77 mus musculu
5	221.5	14.0	359	4	Q86VZ1	Q86vz1 homo sapien
6	209	13.2	319	4	Q8N580	Q8n580 homo sapien
7	207	13.1	343	13	Q8JZ23	Q8jz23 oncorhynchu
8	206.5	13.0	365	11	Q8BUD0	Q8bud0 mus musculu
9	203.5	12.8	367	6	Q8MIO4	Q8mio4 ovis aries
10	200	12.6	357	11	Q7TMV7	Q7tmv7 mus musculu
11	199	12.6	296	11	Q9WU09	Q9wu09 rattus norv
12	198	12.5	339	4	Q8N587	Q8n587 homo sapien
13	198	12.5	358	11	Q8BYC4	Q8byc4 mus musculu
14	198	12.5	376	4	Q7Z3W3	Q7z3w3 homo sapien
15	197.5	12.5	369	11	Q8BJB7	Q8bjb7 mus musculu
16	197	12.4	247	13	Q7T1B7	Q7t1e7 oncorhynchu
17	196	12.4	359	13	Q90VY4	Q90vy4 oncorhynchu
18	194.5	12.3	350	4	Q8N6T6	Q8n6t6 homo sapien
19	193.5	12.2	353	11	O35797	O35797 rattus norv
20	193	12.2	347	13	Q7ZZA4	Q7zza4 brachydanio
21	193	12.2	385	11	Q9JK40	Q9jk40 mus musculu
22	191	12.1	359	11	Q90VZ2	Q90vz2 oncorhynchu
23	191	12.1	377	13	Q7T2S9	Q7t2s9 oncorhynchu
24	190	12.0	390	13	Q8AXM7	Q8axm7 carassius a
25	188	11.9	319	4	Q9Y2T6	Q9y2t6 homo sapien
26	188	11.9	355	6	Q8HZN7	Q8hzn7 gorilla gor
27	187.5	11.8	351	6	Q9MYJ9	Q9myj9 oryctolagus
28	187	11.8	359	13	Q90ZZ2	Q90zz2 oncorhynchu
29	186	11.7	355	6	Q8HZN5	Q8hzn5 macaca mula
30	186	11.7	380	13	Q9DGQ6	Q9dgq6 carassius a
31	185.5	11.7	319	4	Q9NQ20	Q9nq20 homo sapien
32	185	11.7	355	6	Q8HZN4	Q8hzn4 cercopithec
33	184.5	11.6	352	11	Q810T4	Q810t4 cavia porce
34	184.5	11.6	355	6	Q8BVW4	Q8bvw4 mus musculu
35	184	11.6	355	6	Q8HZN6	Q8hzn6 pongo pygma
36	184	11.6	359	6	Q8HZN3	Q8hzn3 papio hamad
37	184	11.6	359	13	Q90ZZ1	Q90zz1 oncorhynchu
38	184	11.6	359	13	Q90ZZ3	Q90zz3 oncorhynchu
39	184	11.6	366	6	Q867B2	Q867b2 capra hircu
40	183.5	11.6	370	13	Q8UWL5	Q8uwl5 fugu rubrip
41	183	11.6	351	11	Q810W6	Q810w6 mus musculu
42	182.5	11.5	355	4	Q8IUZ1	Q8iuz1 homo sapien
43	182.5	11.5	355	6	Q8HZN8	Q8hzn8 pan troglod
44	182.5	11.5	355	11	Q8BMH9	Q8bmh9 mus musculu
45	182.5	11.5	359	13	Q9PVY7	Q9pvy7 anguilla an

ALIGNMENTS

RESULT 1
Q8K3T5
ID   Q8K3T5   PRELIMINARY;   PRT;   300 AA.
AC   Q8K3T5;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   GPR40.
GN   GPR40.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c;
RA   Andrews J.L., Briscoe C.P., Ignar D.M., Muir A.I., Sauls H.R. Jr.,
RA   Tadayyon M.;
RT   "Method of Screening for GPR40 Ligands.";
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c;
RA   Ellis C., Elshourbagy N.;
RL   Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF539809; AAN03478.1; -.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR   GO; GO:0007186; P:G-protein coupled receptor protein signalin...; IEA.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
SQ   SEQUENCE 300 AA; 31837 MW; AFF180F1686F8C17 CRC64;

Query Match   83.9%; Score 1329; DB 11; Length 300;
Best Local Similarity   82.7%; Pred. No. 4.2e-99;
Matches 248; Conservative   17; Mismatches   35; Indels   0; Gaps   0;

```
Qy    1 MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVVALNLGCSDLLLTVSLP  60
Db    1 MDLPPQFSFALYVSAFALGFPLNLLAIRGAVSHAKLRLTPSLVYTLHLGCSDLLLAITLP  60

Qy   61 LKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPPLGYQAFRRP 120
Db   61 LKAVEALASGAWPLPLPFCPVFALAHFAPLYAGGGFLAALSAGRYLGAAFPFGYQAIRRP  120

Qy  121 CYSWGVCAAIWALVLCHLGLVFGLEAPGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG  180
Db  121 RYSWGVCAIWALVLCHLGLALGLETSGSWLDNSTSSLGINIPVNGSPVCLEAWDPDSAR  180

Qy  181 PARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY  240
Db  181 PARLSFSILLFFLPLVITAFCYVGCLRALVRSGLSHKRKLRAAWVAGGALLTLLLCLGPY  240

Qy  241 NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK 300
Db  241 NASNVASFINPDLGGSWRKLGLITGAWSVVLNPLVTGYLGTGPGRGTICVTRTQRGTIQK 300
```

```
RESULT 2
Q8K3T4
ID   Q8K3T4        PRELIMINARY;       PRT;   300 AA.
AC   Q8K3T4;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   GPR40.
GN   GPR40.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley;
RA   Ellis C.E., Elshourbagy N.;
RT   "Rat GPR40 cDNA.";   -
RL   Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF539810; AAN03479.1; -.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR   GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; IEA.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
SQ   SEQUENCE   300 AA;  31835 MW;  4E882A8259A5DD4F CRC64;

Query Match             83.3%;  Score 1320;  DB 11;  Length 300;
Best Local Similarity   81.7%;  Pred. No. 2.2e-98;
Matches  245;  Conservative  19;  Mismatches   36;  Indels    0;  Gaps    0;
```

```
Qy    1 MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVVALNLGCSDLLLTVSLP  60
Db    1 MDLPPQLSFALYVSAFALGFPLNLLAIRGAVSHAKLRLTPSLVVTLHLACSDLLLAITLP  60

Qy   61 LKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPPLGYQAFRRP 120
Db   61 LKAVEALASGAWPLPLPFCPVFALAHFAPLYAGGGFLAALSAGRYLGAAFPFGYQAIRRP  120

Qy  121 CYSWGVCAAIWALVLCHLGLVFGLEAPGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG  180
Db  121 CYSWGVCAIWALVLCHLGLALGLEAPRGWDNTTSSLGINIPVNGSPVCLEAWDPDSAR  180

Qy  181 PARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY  240
Db  181 PARLSFSILLFFLPLVITAFCYVGCLRALVHSGLSHKRKLRAAWVAGGALLTLLLCLGPY  240

Qy  241 NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK 300
Db  241 NASNVASFINPDLEGSWRKLGLITGAWSVVLNPLVTGYLGTGPGQGTICVTRTPRGTIQK 300
```

```
RESULT 3
Q8VCK6
ID   Q8VCK6        PRELIMINARY;       PRT;   330 AA.
AC   Q8VCK6;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Similar to G protein-coupled receptor 43 (Orphan GPCR protein).
GN   GPR43 OR LSSIG.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon;
RA   Strausberg R.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SL;
RA   Senga T., Iwamoto S., Yoshida T., Yokota T., Adachi K., Azuma E.,
RA   Hamaguchi M., Iwamoto T.;
RT   "LSSIG is a novel murine leukocyte specific GPCR that is induced by
RT   the activation of STAT3.";
RL   Blood 0:0-0(2003).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Colon;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; BC019570; AAH19570.1; -.
DR   EMBL; AF545043; AAO16236.1; -.
DR   EMBL; AK078861; BAC37425.1; -.
     MGD; MGI:2441731; Gpr43.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0004872; F:receptor activity; IEA.
DR   GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR   GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; IEA.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   Receptor.
SQ   SEQUENCE   330 AA;  37123 MW;  7AB16802561B8B46 CRC64;

Query Match             21.5%;  Score 340.5;  DB 11;  Length 330;
Best Local Similarity   30.9%;  Pred. No. 1.6e-19;
Matches   99;  Conservative  45;  Mismatches  129;  Indels   47;  Gaps   11;
```

```
Qy   12 YVAAFALGFPLNVLAIRGATAHARL-RLTPSLVYALNLGCSDLLLTVSLPLKAVEALASG  70
Db   14 YILIFLTGLPANLLALRAFMGRVRQPQPAPVHILLNLTLADLLLLLLLPFRIVEAASNF   73

Qy   71 AWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPPLGYQAFRRPCYSWGVCAAI 130
Db   74 RWYLPKIVCALTGFGFYSSIICSTWLLAGISMERYLGVAFPVQYKLSRRPLY--GVIAAL  131

Qy  131 --WALVLCHLGLVFGLEAPGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG---PARFS  185
Db  132 VAWIMSFGHCTIVIVQ------YLNSTEQVGT--ENQITCYENFTQEQLDVVLPVRLE   182

Qy  186 LSLLLFFLPLAITAFCY---VGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPYNA  242
Db  183 LCLVLFFVPMAVTIFCYWRFWIMLTQPHVGAQRRR--RAVGLAVVTLLNFLVCFGPYNM  240
```

Qy   243 SNVASFLYPNLGGSWRKLGLITGAWSVVLNPLV----------------TGYL 279
          |::: |||| :: : :|:|:|:|
Db   241 SHLVGP-YLRQSPSWRVEAVVFSSLNASLDPLLFYFSSSVVRRAFGKGLLLIRNPASSML 299

Qy   280 GRGPGLKTVCAARTQGGKSQ 299
          || || | ||||:|
Db   300 GRG-AKETVEGTKMDRGGSQ 318

RESULT 4
Q8BZ77
ID   Q8BZ77;   PRELIMINARY;   PRT;   396 AA.
AC   Q8BZ77;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Protease-activated receptor 4.
GN   F2RL3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Bone;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium;
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK036427; BAC29423.1; -.
DR   MGD; MGI:1298207; F2rl3.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR   GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; IEA.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
SQ   SEQUENCE   396 AA; 42785 MW; DCI05O2E7AAE1B86 CRC64;

Query Match   14.0%; Score 222.5; DB 11; Length 396;
Best Local Similarity   31.1%; Pred. No. 5.9e-10;
Matches   88; Conservative   27; Mismatches 131; Indels 37; Gaps 10;

Qy   3 LPPQLSFGLYVAARFALGFPLNVLAIRG-ATAHARLRLTPSLVYALNLGCSDLLLTVSLPL 61
          :|: ||||| :|||||| |: || : :|| ||||: ||||:| |||||||||
Db   87 VPTRLVPALYGLVVAVGLPANGLALWVLATRVPRL---PSTILLMNLAVADLLLALVLPP 143

Qy   62 KAVEBALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAPPLGYQAFRRPC 121
          : | | ||||| |||||| |:|| |||:|||||||||||| ||||:|
Db   144 RLAYHLRGGRWPFGEAACRVATARALYGHMYGSVLLLAAVSLDRYLALVHPLRARALRGQR 203

Qy   122 YSWGVCAAIW------ALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWD 175
          || ||| :|| | || || ||:|||||||||||| |||:| :|
Db   204 LTTGLCLVAWLSAATLALPLTLHRQTFRLAGSDRMLCHDALPLTEQT-------SHWR 254

Qy   176 PASAGPARFSLSLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLL 235
          | || || |||||||| || | | |||||||:|||||:|||| ||| ||| |||||
Db   255 PAF------ICLAVLGCFVPLLAMGLCYGATLRALAAANGQRYSHALRL-----TALVLFS 303

Qy   236 CVGPYNASNVASFL-YPNLG-GSWRKLGLITGAW--SVVLNPL 274
          | || | | || :||| | :|: |:| ||:|
Db   304 AVASFTPSNVLLVLHYSNSPEAW---GNLYGAYVPSLALSTL 343

RESULT 5
Q86VZ1
ID   Q86VZ1;   PRELIMINARY;   PRT;   359 AA.
AC   Q86VZ1;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lymph;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lymph;
RA   Strausberg R.;
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC043610; AAH43610.1; -.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR   GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; IEA.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   359 AA; 40635 MW; 565C43660B3C0CF7 CRC64;

Query Match   14.0%; Score 221.5; DB 4; Length 359;
Best Local Similarity   27.4%; Pred. No. 6.4e-10;
Matches   78; Conservative   47; Mismatches 131; Indels 29; Gaps 10;

Qy   11 LYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLPLKAVEBALASG 70
          ||||| || |||||| :| | :|||||||||||||||| |||||||| |
Db   29 VYSLVAAVSIPGNLFSLWVLCRRMGPR-SPSVIFMINLSVTDLMLASVLPFQIYYHCNRH 87

Qy   71 AWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAPPLGYQAFRRPCYSWGVCAAI 130
          | || | | ||:||| :| ||| ||||:|||||| :|
Db   88 HWVFGVLLCNVVTVAFYANMYSSILTMTCISVERFLGVLYPLSSKRWRRRYAVAACAGT 147

Qy   131 WALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEA--WD--PASAGPA--RF 184
          || |||| ||||| | ||| :|| || | :||: :|||
Db   148 WLLLLTALSPLARTD-----------LTYPVHALGIIT------CFDVLKWTMLPSVAMWAVFLF 195

Qy   185 SLSLLLFFLPLAITAFCYVGCLRALARSGLTHRR--KLRAAWVAGGALLTLLCVGPYN- 241
          || |||| | |:| ||||| ||| |||| :||| ||| ||:||| |||| |||
Db   196 TIFILLFLIPFVITVACYTATILKLLRTEEAHGREQRRAVGLAAVVLLAFVTCFAPNNF 255

Qy   242 --ASNVASFLYPNLGGSW---RKLGLITGAWSVLNPLVTGYLGR 281
          ||| ||||| |||| :|: :||||:| ||| ||| ||||
Db   256 VLLAHIVSRLF--YGKSYYHVYKLTLCLSCLNNCLDPFVYFASR 298

RESULT 6
Q8N580
ID   Q8N580;   PRELIMINARY;   PRT;   319 AA.
AC   Q8N580;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to G protein-coupled receptor 55.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RA   Strausberg R.;
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC032694; AAH32694.1; -.
DR   Genew; HGNC:4511; GPR55.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0004872; F:receptor activity; IEA.
DR   GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR   GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; IEA.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   Receptor.
SQ   SEQUENCE   319 AA;   36637 MW;   D6E5C6CA8426E7D5 CRC64;

Query Match   13.2%;   Score 209;   DB 4;   Length 319;
Best Local Similarity   27.4%;   Pred. No. 5.8e-09;
Matches   73;   Conservative   33;   Mismatches 110;   Indels   50;   Gaps   9;

Qy   2   DLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLR---LTPSLVYALNLGCSDLLLTVS 58
                :||||||||||||||| |||| :: :|||||| : :: :||| ||||||
Db   17   ELMKTLQPAVHIPTFVLGLLLNLLAIHGFSTFLKNRWPDYAATSIYMINLAVEFDLLLVLS 76

Qy   59   LPLKAVEALASGAWPLPASLCPVFAVAHFPPLYAGGFLAALSAGRYLGAAFPLGYQAFR 118
           ||| :: |||||| |||||||||||||||:||||||||
Db   77   LPFKMVLSQVQSPPP---SLCTLVECLYFVSMYGSVFTICFISMDRFLAIRYPLLVSHLR 133

Qy   119 RPCYSWGVCAATWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAW---- 174
           : : |||||| ||||:|||| | |: : ||:|||||
Db   134 SPRKIFGICCTIWVLV------------------WTG------SIPIYSFHGKVEKYMCFH 170

Qy   175 ---DPASAGPARFSLSLLLFFLPLAITTAFCYVGCLRALARSGLTHRRKLRAAWVAGGALL 231
              |||:||||||||| ||||| |:: ||||||| : :||||||||||||||
Db   171 NMSDDTWSAKVFPFPLEVPGFLLMGIMGFCC---CSRSI--HILLGRRDHTQDWVQQKA-- 223

Qy   232 TLLLCVGPYNAS--NVASFLYPNLG 254
                ||||||||| :|||| ||
Db   224 ----CIYSIAASLAVFVVSFLPVHLG 245

RESULT 7
Q8JJ23
ID   Q8JJ23;   PRELIMINARY;   PRT;   343 AA.
AC   Q8JJ23;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   G protein-coupled receptor rhodopsin-1.
OS   Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC   Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX   NCBI_TaxID=8022;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kiiryu I., Koellner B., Kuroda A., Ototake M., Dijkstra J.M.;
RT   "A new probable G-protein coupled receptor gene associated with the
RT   immune system of rainbow trout (Oncorhynchus mykiss).";
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF472607; AAM21607.1; -.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0004872; F:receptor activity; IEA.
DR   GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR   GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; IEA.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   Receptor.
SQ   SEQUENCE   343 AA;   38830 MW;   761F06E57F76BB40 CRC64;

Query Match   13.1%;   Score 207;   DB 13;   Length 343;
Best Local Similarity   23.7%;   Pred. No. 9e-09;
Matches   76;   Conservative   55;   Mismatches 124;   Indels   66;   Gaps   10;

Qy   1   MDLPPQLSFG----------------------LYVAAFALGFPLNVLAIRGATAHARLR 37
           ||:||
Db   1   MDMTPTPSSGNSNLNEANNCGVDLSQDAIYLPVIYSIFFIIGTPLNIMALFGL-----YR 55

Qy   38   LTPS----LVYALNLGCSDLLLTVSLPLKAVEEALASGAWPLPASLCPVFAVAHFPPLYAG 93
               :   :|| || :| :|||||||||||||||||||||||||||||||||||||||
Db   56   LIKSENVLPVYVINLLLSDLLQLFTVPLWIDYYRRGHSWRFGSTSCQLLGVSFYISITYG 115

Qy   94   GGFLAALSAGRYLGAAFPLGYQAFRRPCYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDH 153
           :|: ||| ||| | ||||:| |: :: | :||| :|| :: :
Db   116 IAFMCIIALERYLAIAKPLRFQALRKLKPARWIALSIWVVVAVPPSIV----------- 163

Qy   154 SNTSLGINTPVNGSPVCLEAWDPASAGPARFSLSL--LLFFLPLAITTAFCYVGCLR---A 208
             ::||:::   :: || :|||| ||| | |:||||| |:::|:
Db   164 ----LHKMQPNDNHTLCIESY-PSKEGFIYRLITLSLSFIIPLAFVIILHRKTLRSLSA 218

Qy   209 LARSGLTHRRKLRAAWVAGGALLTLLLCVGPYNASNVASFLYPNLGG---SWRK------ 259
           |||| ||| |:: : | ::|:|||| |:|::|| |||||
Db   219 IGTLGTEEKHIRGLLIL--LMVIFILVLGPYHITGCVKYIGLLLHGDACEWEKTVFVPY 276

Qy   260 -LGLITGAWSVVLNPLVTGYL 279
            | | : : |||:|: |
Db   277 QLGRGLLSLNSVLDFILYTFL 297

RESULT 8
Q8BUD0
ID   Q8BUD0;   PRELIMINARY;   PRT;   365 AA.
AC   Q8BUD0;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Probable G protein-coupled receptor GPR4.
GN   GPR4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Heart;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium;
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK085829; BAC39547.1; -.
DR   MGD; MGI:2441992; Gpr4.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR   GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; IEA.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.

```
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
SQ   SEQUENCE   365 AA;  41104 MW;  2D9F8128DE0EA37 CRC64;

Query Match             13.0%;
Best Local Similarity   27.5%;   Score 206.5; DB 11;  Pred. No. 1e-08;
Matches  75;  Conservative  35;  Mismatches 108;  Indels  55;  Gaps  11;

Qy    4 PPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSL-VYAINLGCSDLLLTVSLPLK  62
Db   22 PP----SLYIFVIGVGLPTNCLAL--WAAYRQVRQHNELGVYLMNLSIADLLYICTLPLW  75

Qy   63 AVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRPCY 122
Db   76 VDYFPLHHDNWIHGPGSCKLFGFIFYSNIYISIAFLCCISVDRYLAVAHPLRFARLRRVKT 135

Qy  123 SWGVCAAIWALVLCHLGLVFGLRAPGGWLDHSNTSLGINTPVNGSPV------------ 169
Db  136 AVAVSSVVNA------------------TELG---ANSAPLFHDELFRDRYNHT       168

Qy  170 -CLEAWDPASAGPARFSLSLLL--FFLPLAITAPCYVGCLRALARSGLTHRR-KLRAAWV 225
Db  169 FCFEKF-PMERWVAWMNLYRVFVGFLFPWALMLLCYRGLIRAVQSSVSTERQEKVKIRL  227

Qy  226 AGGALLTLLLCVGPYNASNVA-SFLYPNLGGSW 257
Db  228 ALSLIAIVLVCFAPYHALLLSRSAVY--LGRPW 258
```

RESULT 9
Q8MI04

```
ID   Q8MI04;      PRELIMINARY;       PRT;   367 AA.
AC   Q8MI04;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Somatostatin receptor subtype 5.
GN   SST5.
OS   Ovis aries (Sheep).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Caprinae; Ovis.
OX   NCBI_TaxID=9940;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pituitary;
RA   Debus N., Dutour A., Boudouresque F., Vuaroqueaux V., Oliver C.,
RA   Ouafik L.;
RT   "Molecular cloning and tissue distribution of the ovine somatostatin
RT   receptor subtype 5: osst5.";
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pituitary;
RA   Debus N.;
RT   "Regulation of GH secretion in sheep: Role of somatostatin and its
RT   receptors.";
RL   Thesis (1999).
RL   Department of Intercellular communications in Endocrinology,
RL   University of Aix-Marseilles II, Marseilles, France.
DR   EMBL; AJ441116; CAD29615.1; -.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0004872; F:receptor activity; IEA.
DR   GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR   GO; GO:0007186; P:G-protein coupled receptor protein signalin. ..; IEA.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   Receptor.
FT   CHAIN         2    367       SOMATOSTATIN RECEPTOR SUBTYPE 5.
```

```
SQ   SEQUENCE   367 AA;  40269 MW;  FA9CAB7AA8BBD764 CRC64;

Query Match             12.8%;   Score 203.5; DB 6;   Length 367;
Best Local Similarity   26.9%;             Pred. No. 1.8e-08;
Matches  84;  Conservative  48;  Mismatches 131;  Indels  49;  Gaps  15;

Qy   11 LYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLPLKAVERALASG  70
Db   46 LYLLVCAVGLGGNALVIYVVLRHAKMK-TVTNIYILNLAVADVLLMLGLPFVATQN-AIS 103

Qy   71 AWPLPASLCPVF----AVAHFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRPCYSWGV  126
Db  104 YWPFGPVLCRLVMTLDGINQFTSIFC---LTVMSVDRYLAVHPIRSARWRRPRVAKLA  159

Qy  127 CAAIWAL-VLCHLGLVFGLRAPGGWLDHSNTSLGINTPVNGSPVCLEAW-DPASAGPARF 184
Db  160 SAAVWAFSLVMSLPLVVFADIQEGW------NT------CNLTWPEVGLWGAVF       202

Qy  185 SL--SLLLFFLPLAITAFCYVGCLRALARSGL----THRRKLRAAWVAGGALLTLLLCVG 238
Db  203 IIYTSVLGFPGPLLVICLCVVLIVVKLKASGVRVGSTRRSERK--VTRMVVVVVLVFAG  260

Qy  239 ---PYNASNVASFLY-----PNLGGSWRKLGLITGAWSVVLNPLVTGYLG---RGPGLKT 287
Db  261 CWLPFFIVNIVNLAFALPEEPASAGLYFFVVLSYANSCA-NPCLYGFLSDNLRQSFRKV  319

Qy  288 VCAARTQGGKSQ 299
Db  320 LCLRKGYGAGAE 331
```

RESULT 10
Q7TMV7

```
ID   Q7TMV7;      PRELIMINARY;       PRT;   357 AA.
AC   Q7TMV7;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6NCr; TISSUE=Hematopoietic Stem Cell;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6NCr; TISSUE=Hematopoietic Stem Cell;
RA   Strausberg R.;
RL   Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC052868; AAH52868.1; -.
```

```
KW    Hypothetical protein.
SQ    SEQUENCE  357 AA;  40191 MW;  56B96775303FBDDF CRC64;

Query Match           12.6%;  Score 200;  DB 11;  Length 357;
Best Local Similarity 25.8%;  Pred. No. 3.4e-08;
Matches     65;  Conservative   38;  Mismatches  103;  Indels  46;  Gaps   9;

Qy     12  YVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLPLKAVEALASGA  71
               :|| | :||:  |:|| || : |::|  :|
Db     34  YSLVFIIGLVGNLLAL-VVIVQNRKKINSTTLYSMNLVISDILFTTALPTRIAYYALGFD  92

Qy     72  WPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRPCYSWGVCAAIW  131
               :|| :|   ||   |:|| |||| ||| ::||| : |:| | || : : |
Db     93  WRIGDALCRVTALVFYINTYAGVNFMTCLSIDRFPAVVHPLRYNKIRIEYAKGVCLSVW  152

Qy    132  ALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVN----GSPVCLEAWDPASAGPAR----  183
                                                          :|:
Db    153  ILVFAQTLPLF------------------LTPMSKEEGDKTTCMEY--PNFEGTASLPWI  192

Qy    184  -FSLSLLLFFLPLAITAFCY--VGC--LRALARSGLTHRRKLRAAWVAGGALLTL-----  233
            |  :|   |  | :|  |                 |:|  :|:|
Db    193  LLGACLLGGVVLPITVILLCYSQICKLPRTAKQNPLTEK-----SGVNKKALNTIILLIV  247

Qy    234  --LLCVGPYNAS  243
              :||
Db    248  VFILCFFTPYHVA  259


RESULT 11
Q9WU09
ID    Q9WU09         PRELIMINARY;        PRT;    296 AA.
AC    Q9WU09;
DT    01-NOV-1999  (TrEMBLrel. 12, Created)
DT    01-NOV-1999  (TrEMBLrel. 12, Last sequence update)
DT    01-JUN-2003  (TrEMBLrel. 24, Last annotation update)
DE    G protein-coupled receptor (Fragment).
GN    GPR55.
OS    Rattus norvegicus (Rat).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX    NCBI_TaxID=10116;
RN    [1]
RP    SEQUENCE FROM N.A.
RX    MEDLINE=99132385; PubMed=9931487;
RA    Sawdzargo M., Nguyen T., Lee D.K., Lynch K.R., Cheng R., Heng H.H.Q.,
RA    George S.R., O'Dowd B.F.;
RT    "Identification and cloning of three novel human G protein-coupled
RT    receptor genes GPR52, PSI.GPR53 and GPR55: GPR55 is extensively
RT    expressed in human brain.";
RL    Mol. Brain Res. 64:193-198(1999).
DR    EMBL; AF100789; AAD22411.1; -.
DR    GO; GO:0016021; C:integral to membrane; IEA.
DR    GO; GO:0004872; F:receptor activity; IEA.
DR    GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR    GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; IEA.
DR    InterPro; IPR000276; GPCR_Rhodpsn.
DR    Pfam; PF00001; 7tm_1; 1.
DR    PRINTS; PR00237; GPCRRHODOPSN.
DR    PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW    Receptor.
FT    NON_TER        1      1
FT    NON_TER      296    296
SQ    SEQUENCE  296 AA;  34008 MW;  7D7A9596C21A030 CRC64;

Query Match           12.6%;  Score 199;  DB 11;  Length 296;
Best Local Similarity 25.8%;  Pred. No. 3.4e-08;
Matches     67;  Conservative   34;  Mismatches  129;  Indels  30;  Gaps   6;

Qy      2  DLPPQLSFGLYVAAFALGFPLNVLAIRGATA--HARLRLTPSLVYALNLGCSDLLLTVS  58
            :| :|     :|| | :||:  |:|| || : |::|  :|
Db     19  NLRKTLQLAVHIPTFLLGLVLNLLAIRGFGAFLKRRQLEYMATSIYMINLACFDLLLVLS  78

Qy     59  LPLKAVEALASGAWMPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFR  118
```

```
KW    |||| ::                :|| |  ::
SQ    LPFKRMLPQVKSSPSL--VFCTFVECLYFISMYGSVFTICFISLDRLLAIQYPLLVNHFR  136

Db     79

Qy    119  RPCYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINT---PVNGSPVCLEAWD  175
                :::||:::|||          | | :|              :| |
Db    137  SPRKTFGICCIIWMLV--------------------WVG---SIPIYTFHRGVEGYKCFHNMSD  177

Qy    176  PASAGPARFSLSLLLFFLPLAITAFC---YVGCLRALARSGLTHRRKLRAAWVAGGALLT  232
               |::                  |  :|::                 :::
Db    178  STWSARVFFPLEIFGFLLPMGIMGFCSYRSIHILLSIQGDDQWVRKRACIIWTIATNLVV  237

Qy    233  LLCVGPYNASNVASFLYPN  252
           :::    |:
Db    238  FVVSFLPVHLGLFLQFLVRN  257


RESULT 12
Q8N5S7
ID    Q8N5S7         PRELIMINARY;        PRT;    339 AA.
AC    Q8N5S7;
DT    01-OCT-2002  (TrEMBLrel. 22, Created)
DT    01-OCT-2002  (TrEMBLrel. 22, Last sequence update)
DT    01-JUN-2003  (TrEMBLrel. 24, Last annotation update)
DE    G protein-coupled receptor 17.
OS    Homo sapiens (Human).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX    NCBI_TaxID=9606;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    TISSUE=Brain, and Lung;
RA    Strausberg R.;
RL    Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
DR    EMBL; BC031653; AAH31653.1; -.
DR    GO; GO:0016021; C:integral to membrane; IEA.
DR    GO; GO:0004872; F:receptor activity; IEA.
DR    GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR    GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; IEA.
DR    InterPro; IPR000276; GPCR_Rhodpsn.
DR    Pfam; PF00001; 7tm_1; 1.
DR    PRINTS; PR00237; GPCRRHODOPSN.
DR    PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR    PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW    Receptor.
SQ    SEQUENCE  339 AA;  37860 MW;  BB0CAFD0FC371D63 CRC64;

Query Match           12.5%;  Score 198;  DB 4;  Length 339;
Best Local Similarity 26.6%;  Pred. No. 4.7e-08;
Matches     69;  Conservative   41;  Mismatches  113;  Indels  36;  Gaps   9;

Qy     11  LYVAAFALGFPLNVLAIRGATAHARLRL-----------TPSLVYALNLGCSDLLLTVSLP  60
              |||: ||: | ||    :: :::                :|| |  ::
Db     33  LFASFYLLDF---ILALVGNT-----LALWFFIRDHKSGTPANVFLMHLAVADLSCVLVLP  85

Qy     61  LKAVEALASGAWMPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRP  120
            |:                            |  :|| |||| ||| ::||| : |:| | |
Db     86  TRLVYHFSGNHWPFGEIACRLTGFLFYLNMYASIYFLTCISADRFLAIVHPVKSLKLRRP  145

Qy    121  CYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG  180
               :::||:::|||          | | :|            :|
Db    146  LYAHLACAFLWVVVAVAMAPL--LVSPQ----------------TVQTNHTVVCLQLYREKASH  191

Qy    181  PARFSLSLLLFFLPLAITAFCYVGCLRALARSGL--THRRKLRAAWVAGGALLTLLLCVG  238
            |::                 |  :|::            :::
Db    192  HALVSLA-VAFTFPITTVTCYLLIIRSL-RQGLRVEKRLKTKAVRMIAIVLAIFIVCFV  249

Qy    239  PYNASNVASFL-YPNLGGS  256
           :::        |:
Db    250  PYHVNRSVVLHYRSHGAS  268


RESULT 13
Q8BYC4
```

```
ID   Q8BYC4       PRELIMINARY;    PRT;   358 AA.
AC   Q8BYC4;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Putative G-protein coupled receptor homolog.
GN   GPR20 OR A430106B11RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Thymus;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573 (2002)
DR   EMBL; AK040547; BAC30623.1; -.
DR   MGD; MGI:2441803; Gpr20.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR   GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; IEA.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
SQ   SEQUENCE   358 AA;   39431 MW;   C6F7842ABDB9CA8F CRC64;

Query Match             12.5%;   Score 198;  DB 11;   Length 358;
Best Local Similarity   28.0%;   Pred. No. 5e-08;
Matches    69;  Conservative  41;  Mismatches  90;  Indels  46;  Gaps   9;

Qy     16  FALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLPLK-AVEALASG----   70
               : | ||:||:|| |:  : ||: :|   ::|:|  |: ::|:||
Db     62  FLAGLVLNGLALYVFCCRTRAK-TPSVTYTINLVVTDLLVGLSLPTRFAVFYGARGCLRC  120

Qy     71  AWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAPPLGYQAFRRPCYSWGVCAAI  130
               : |      ||: :|  :|:: : |    |  ||||    |      |:
Db    121  AFP--------HVLGYFLNMHCSILFLTCICVDRYLAIVQPEGSRRWRQPACAKAVCIFV  172

Qy    131  WALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAGPARFSLSLLL  190
               |:|        :|:                  :|:| :
Db    173  WLAA-------GVV--------------TISVLGYK---SGGRSCCRV-------FALTVLE  203

Qy    191  FFLPLAITAFCYVGCLRALARSGLT--HRRKLRAAWVAGGALLTLLCVGPYNASNVAS  247
               | |:|:::: |    :|    ||:   ||:: :|  ::|:: |::::|:
Db    204  FLLPLLVISVFTGRIMCALSRPGLLRQGRQRVRAMQLLLTVLVIFLVCFTPFHARQVAV  263

Qy    248  FLYPNL  253
Db    264  ALWPNV  269


RESULT 14
Q7Z3W3
ID   Q7Z3W3       PRELIMINARY;    PRT;   376 AA.
AC   Q7Z3W3;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein DKFZp686N1782.
GN   DKFZp686N1782.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Human retina;
RA   Wambutt R., Heubner D., Mewes H.W., Weil B., Amid C., Osanger A.,
RA   Fobo G., Han M., Wiemann S.;
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BX537386; CAD97628.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   376 AA;   42720 MW;   7638373C15A695DD CRC64;

Query Match             12.5%;   Score 198;  DB 4;   Length 376;
Best Local Similarity   24.5%;   Pred. No. 5.2e-08;
Matches    73;  Conservative  42;  Mismatches 145;  Indels  38;  Gaps   9;

Qy      3  LPPQLSFGLYAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLPLK   62
              :  |: |||||  |: ||: |: |:  :| :: :||     :|:|:  ||:|  :
Db     92  LSTKLIPAIYLLVFVVGVPANAVTL--WMLFPRTRSICTTVFYTNLAIADFLFCVTLPFK  149

Qy     63  AVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAPPLGYQAFRRPCY  122
               : |:: ||  |:|: :||||  |:: || :|  |:   |  |:||     |  |: |:
Db    150  IAYHLNGNNWVFGEVLCRATTVIFYGNMYCSILLACISINRYLAIVHPFTYRGLPKHTY  209

Qy    123  SWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLG--INTPVNGSPVCLEAWDPASAG  180
               :  |  ||::|::|  :|::    :   |:          |::  |::     |    |::
Db    210  ALVTCGLVWATVFLVMLPFFILKQBYYLVQPDITTCHDVHNTCESSSPFQLYYF------  263

Qy    181  PARFSLSLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLL--LCV  237
               |  ||:||:||  |:||: || |: : |  ::|  :||:||||    :|:||    |:|
Db    264  ---ISLAFFGFLIPFVLIIYCYAAIIRTL--NAYDH---RWLWYVKASLLILVIFTICF  314

Qy    238  GPYNASNV---ASFLYPNLGGSW--RKLGLITGAWSVVLNPLV-------------TGYL  279
               .::::: :   ||.: :| :  :     |  || |:: ::  ::             |::|
Db    315  APSNILIIHHANYYNNTDGLYFIYLIALCLGSLNSCLDPFLYFLMSKTRNHSTAYL  372


RESULT 15
Q8BJB7
ID   Q8BJB7       PRELIMINARY;    PRT;   369 AA.
AC   Q8BJB7;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Proteinase activated receptor 3 precursor.
GN   F2RL2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573 (2002)
DR   EMBL; AK089444; BAC40886.1; -.
MGD; MGI:1298208; F2rl2.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR   GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; IEA.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
SQ   SEQUENCE   369 AA;   41696 MW;   C8E5B009710519EF CRC64;

Query Match             12.5%;   Score 197.5;  DB 11;   Length 369;
Best Local Similarity   24.1%;   Pred. No. 5.6e-08;
Matches    70;  Conservative  45;  Mismatches 134;  Indels  41;  Gaps   11;

Qy      3  LPPQLSFGLYAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLPLK   62
              :  |: |||||  |: ||: |: |::   ::|    ::|:|:  ||:| ::
Db     91  LSTQVIPAIYILLFVVGVPANIVTLWKLS-LRTKSISLVIFHTNLAIADLLFCVTLPFK  148

Qy     63  AVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAPPLGYQAFRRPCY  122
```

```
Db  149 IAYHLNGNNWVFGEVTCRITTVVFYGNMYCAILILTCMGINRYLATAHPFTYQKLPKRSF 208
            |   :    |          :      |   :  |         |||    | |||    ||     :
Qy  123 SWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAGPA 182
            :  :  :|: :: |:|:   |  |    :       |||        :     ||    :
Db  209 SMLMCGMVWMVFLYM-LPFVILKQEYHLVHSEIT----------TCHDVVD-ACBSPS 255

Qy  183 RF------SLSLLFFLPLAITAFCYVGCLRALARSGLTHRRKLR-AAWVA--GGALLTL 233
            |:|     |::  |  |:|          |||  |:|:|  ||      :
Db  256 SFRFYYFVSLAFFGFLIPFVIIIFCY--------TTLIHKLKSKDRIWLGYIKAVLLIL 306

Qy  234 L---LCVGPYNASNV---ASFLYPNLGGSW--RKLGLITGAWSVVLNPLV 275
            :  :|    |     |:::       |  :|  :| :| :|:
Db  307 VIFTICFAPTNIILVIHHANYYYHNTNSLYFMXLIALCLGSLNSCLDPFL 356
```

Search completed: April 29, 2004, 12:11:31
Job time : 47 secs

```
                GenCore version 5.1.6
            Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:   April 29, 2004, 12:04:08 ; Search time 18 Seconds
                                      (without alignments)
                                      867.836 Million cell updates/sec

Title:          US-10-202-687-2
Perfect score:  1584
Sequence:       1 MDLPPQLSFGLYVAARFALGP.........RGPGLKTVCAARTQGGKSQK 300

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters:        141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID		Description
1	1584	100.0	300	1	GP40_HUMAN	O14842	homo sapien
2	366	23.1	346	1	GP42_HUMAN	O15529	homo sapien
3	347	21.9	346	1	GP41_HUMAN	O14843	homo sapien
4	332	21.0	330	1	GP43_HUMAN	O15552	homo sapien
5	244	15.4	420	1	PAR1_XENLA	P47749	xenopus lae
6	231.5	14.6	385	1	PAR4_HUMAN	Q96rio	homo sapien
7	228	14.4	425	1	PAR1_HUMAN	P25116	homo sapien
8	226	14.3	425	1	PAR1_PAPHA	P56488	papio hamad
9	217.5	13.7	372	1	GP92_HUMAN	Q9hico	homo sapien
10	216.5	13.7	358	1	GP20_HUMAN	Q99678	homo sapien
11	213	13.4	428	1	PAR1_CRILO	Q00991	cricetulus
12	208	13.1	328	1	P2Y3_MELGA	O93361	meleagris g
13	208	13.1	430	1	PAR1_MOUSE	P30558	mus musculu
14	205.5	13.0	363	1	SSR5_RAT	P30938	rattus norv
15	204.5	12.9	355	1	IL8A_RABIT	P21109	oryctolagus
16	204	12.9	328	1	P2Y3_CHICK	Q98907	gallus gall
17	204	12.9	432	1	PAR1_RAT	P26824	rattus norv
18	201.5	12.7	356	1	IL8B_CANFA	O97571	canis famil
19	201	12.7	358	1	IL8B_RABIT	P35344	oryctolagus
20	201	12.7	367	1	GP17_HUMAN	Q13304	homo sapien
21	200.5	12.7	396	1	PAR4_MOUSE	O88634	mus musculu
22	198.5	12.5	328	1	P2Y6_HUMAN	Q15077	homo sapien
23	198	12.5	374	1	PAR3_HUMAN	O00254	homo sapien
24	197.5	12.5	328	1	P2Y6_RAT	Q63371	rattus norv
25	197.5	12.5	369	1	PAR3_MOUSE	O08675	mus musculu
26	197.5	12.5	395	1	PAR4_RAT	Q920e0	rattus norv
27	194.5	12.3	350	1	IL8A_HUMAN	P25024	homo sapien
28	194	12.2	364	1	SSR5_HUMAN	P35346	homo sapien
29	193	12.2	362	1	SSR5_MOUSE	O08858	mus musculu
30	191.5	12.1	368	1	PAR3_RAT	Q920e1	rattus norv
31	190.5	12.0	349	1	IL8A_RAT	P70612	rattus norv
32	190.5	12.0	350	1	IL8A_PANTR	P55920	pan troglod
33	190.5	12.0	362	1	GPR4_HUMAN	P46093	homo sapien
34	188.5	11.9	352	1	C5AR_CANFA	P30992	canis famil
35	187	11.8	377	1	P2Y2_HUMAN	P41231	homo sapien
36	186.5	11.8	362	1	GPR4_PIG	P50132	sus scrofa
37	185.5	11.7	319	1	GP31_HUMAN	O00270	homo sapien
38	185.5	11.7	353	1	IL8B_GORGO	Q28422	gorilla gor
39	185	11.7	337	1	C5L2_HUMAN	Q9p296	homo sapien
40	185	11.7	361	1	P2Y4_MOUSE	Q9jj87	mus musculu
41	185	11.7	370	1	OPRX_CAVPO	P47748	cavia porce
42	184.5	11.6	344	1	P2Y5_MOUSE	Q8bmco	mus musculu
43	184.5	11.6	353	1	IL8B_PANTR	Q28807	pan troglod
44	184	11.6	365	1	P2Y4_HUMAN	P51582	homo sapien
45	183.5	11.6	350	1	C5AR_RABIT	Q9tue1	oryctolagus

ALIGNMENTS

```
RESULT 1
GP40_HUMAN
ID   GP40_HUMAN     STANDARD;      PRT;   300 AA.
AC   O14842;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Putative G protein-coupled receptor GPR40.
GN   GPR40.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98008875; PubMed=9344866;
RA   Sawzdargo M., George S.R., Nguyen T., Xu S., Kolakowski L.F. Jr.,
RA   O'Dowd B.F.;
RT   "A cluster of four novel human G protein-coupled receptor genes
RT   occurring in close proximity to CD22 gene on chromosome 19q13.1.";
RL   Biochem. Biophys. Res. Commun. 239:543-547(1997).
CC   -!- FUNCTION: Orphan receptor.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF024687; AAB86710.1; -.
DR   PIR; JC5714; JC5714.
DR   Genew; HGNC:4498; GPR40.
DR   MIM; 603820; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   GO; GO:0004930; F:G-protein coupled receptor activity; TAS.
DR   GO; GO:0007186; P:G-protein coupled receptor protein signalin...; TAS.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; FALSE_NEG.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane.
FT   DOMAIN        1     10       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     11     31       1 (POTENTIAL).
FT   DOMAIN       32     42       CYTOPLASMIC.
FT   TRANSMEM     43     63       2 (POTENTIAL).
FT   DOMAIN       64     81       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     82    102       3 (POTENTIAL).
FT   DOMAIN      103    123       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    124    144       4 (POTENTIAL).
FT   DOMAIN      145    178       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    179    197       5 (POTENTIAL).
```

```
FT   DOMAIN        198    221       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      222    242       6 (POTENTIAL).
FT   DOMAIN        243    258       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      259    279       7 (POTENTIAL).
FT   DOMAIN        280    300       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD      155    155       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE      300 AA;  31457 MW;  77EF27DACD93E80B CRC64;

Query Match
Best Local Similarity   100.0%;   Score 1584;  DB 1;  Length 300;
Matches 300;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1  MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLP   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MDLPPQLSFGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLP   60

Qy   61  LKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRP  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  LKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRP  120

Qy  121  CYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  CYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAG  180

Qy  181  PARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY  240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  PARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPY  240

Qy  241  NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK  300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  NASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRGPGLKTVCAARTQGGKSQK  300

RESULT 2
GP42_HUMAN
ID   GP42_HUMAN     STANDARD;     PRT;    346 AA.
AC   O15529;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Probable G protein-coupled receptor GPR42.
GN   GPR42.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98008875; PubMed=9344866;
RA   Sawzdargo M., George S.R., Nguyen T., Xu S., Kolakowski L.F. Jr.,
RA   O'Dowd B.F.;
RT   "A cluster of four novel human G protein-coupled receptor genes
RT   occurring in close proximity to CD22 gene on chromosome 19q13.1.";
RL   Biochem. Biophys. Res. Commun. 239:543-547(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98008875; PubMed=9344866;
RA   Lamerdin J.E., McCready P.M., Adamson A.W., Burkhart-Schultz K.,
RA   Garcia E., Kyle A., Ramirez M., Stilwagen S., Garnes J., Danganan L.,
RA   Bruce R., Quan G., Montgomery M., Ow D., Kobayashi A., Olsen A.O.,
RA   Carrano A.V.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Orphan receptor.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
```

```
DR   EMBL; AF024689; AAB86712.1; -.
DR   EMBL; AC002511; AAB67885.1; -.
DR   PIR; JC5716; JC5716.
DR   Genew; HGNC:4500; GPR42.
DR   MIM; 603822; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   GO; GO:0004930; F:G-protein coupled receptor activity; TAS.
DR   GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; TAS.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane.
FT   DOMAIN          1     19       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM       20     40       1 (POTENTIAL).
FT   DOMAIN         41     47       CYTOPLASMIC.
FT   TRANSMEM       48     68       2 (POTENTIAL).
FT   DOMAIN         69     90       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM       91    111       3 (POTENTIAL).
FT   DOMAIN        112    132       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      133    153       4 (POTENTIAL).
FT   DOMAIN        154    178       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      179    199       5 (POTENTIAL).
FT   DOMAIN        200    222       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      223    243       6 (POTENTIAL).
FT   DOMAIN        244    258       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      259    279       7 (POTENTIAL).
FT   DOMAIN        280    346       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD      166    166       N-LINKED (GLCNAC. . .). (POTENTIAL).
SQ   SEQUENCE      346 AA;  38695 MW;  03F6742DA4B798D0 CRC64;

Query Match
Best Local Similarity   23.1%;   Score 366;  DB 1;  Length 346;
Matches 101;  Conservative  38;  Mismatches 129;  Indels  28;  Gaps  8;

Qy    1  MDLPPQLS--------FGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCS   51
         |||||||:         ||   |:|  ||:||:|   |::|| |:|
Db    1  MDTGPDQSYFSGNHWFVFSVYLLTPLVGLPLNLLALVVFVGKLRCRPVAVDVLLNLTAS   60

Qy   52  DLLLTVSLPLKAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFP  111
         ||||: ::|||       |  ||| :||| || ||:| ||::  |:|| |:   |:||
Db   61  DLLLLFLFPRMVRAANGMHWPLPFILCPLSGFIFTTIYLTALFLAAVSIERFLSVAHP  120

Qy  112  LGYQAFRRPCYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVC-  170
         | | | |  |  |  |:||  || | |:|  |  |    ||            |     |
Db  121  LWYKTRPRLGQAGLVSVACWLLASAHCSVVYVIEFSGD-ISHSQGTNG-------TCY  170

Qy  171  LEAWDP--ASAGPARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGG  228
         ||  |     :: | |  | |:|   |::||||  ::||:| |  :|:|||: |:
Db  171  LEFWKDQLAILLPVRLEMAVVLFVVPLIITSYCYSRLVWILGRGG--SHRRQRRVAGLVAA  229

Qy  229  ALLTLLLCVGPYNASNVASFLYPNLGGS--WRKLGLITGAWSVVLNPLVTGYLGRG  282
         ||| ||| | |||  :|| :|   |:||  |:|     |   |   |||| : |:|
Db  230  TLLNFLVCFGPYNVSHVVGYI---CGESPVWRIYVTLLSTLNSCVDPFVYFSSSG  282

RESULT 3
GP41_HUMAN
ID   GP41_HUMAN     STANDARD;     PRT;    346 AA.
AC   O14843;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Putative G protein-coupled receptor GPR41.
GN   GPR41.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98008875; PubMed=9344866;
```

```
Query Match     21.9%;  Score 347;  DB 1;  Length 346;
Best Local Similarity  32.3%;   Pred. No. 6e-19;
Matches   95; Conservative  41; Mismatches  134; Indels  24; Gaps  6;

Qy    1 MDLPPQLS------FGLYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCS   51
        ||||||:|        |||:|:| :|||||:|| :.  | || ::|||:.|||||||
Db    1 MDTGPDQSYFSGNHWFVFSVVLLTFLVGLPLNLLALVVFVGKLQRRPVAVDVLLNLTAS  60

Qy   52 DLLLTVSLPLKAVEALASGAWPLPASLCPVFAVAHFPPLYAGGFLAALSAGRYLGAAFP  111
        |||| |||  :|::|. : ||:||| || . .|||:::  ||.||.:||::||:||||
Db   61 DLLLLFLPFRMVEAANGMHWPLPFILCPLSGFIFTTIYLTALFLAAVSIERFLSVAHP  120

Qy  112 LGYQAFRRPCYSWGVCAAIWALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCL  171
        | |: .|| |:   | | :|.:| .  | |.  |. |:  :|||:  :  :|| .   || 
Db  121 LWYKTRPRLGQAGLVSVACWLLASAHCSVVVVIEFSGD-ISHSQGTNG-------TCY  170

Qy  172 EAW---DPASAGPARFSLSLLFFLPLAITAFCYVGCLRALARSGLTHRRKLRAAWVAGG  228
        ::    ||::| ||| || ::  ||:|||:|||. ||::|::|:|:||:::| : : .
Db  171 LEFRKDQLAILLPVRLEMAVVLFVVPLIITSYCYSRLVWILGRGG-SHRRQRRVAGLLAA  229

Qy  229 ALLTLLCVGPYNASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGRG  282
        :||.:| |.||||.:.| :    ||.||.::.| | :.:.: ||:  :  :|
Db  230 TLLNFLVCFGPYNVSHVVGYICGE-SPAWRIYVTLLSTLNSCVDPFVYYFSSSG  282


RESULT 4
GP43_HUMAN
ID   GP43_HUMAN     STANDARD;      PRT;   330 AA.
AC   O15552;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Probable G protein-coupled receptor GPR43.
GN   GPR43.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98008875; PubMed=9344866;
RA   Sawzdargo M., George S.R., Nguyen T., Xu S., Kolakowski L.F. Jr.,
RA   O'Dowd B.F.;
RT   "A cluster of four novel human G protein-coupled receptor genes
RT   occurring in close proximity to CD22 gene on chromosome 19q13.1.";
RL   Biochem. Biophys. Res. Commun. 239:543-547(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Lamerdin J.E., McCready P.M., Adamson A.W., Burkhart-Schultz K.,
RA   Garcia B., Kyle A., Ramirez M., Stilwagen S., Garnes J., Danganan L.,
RA   Bruce R., Quan G., Montgomery M., Ow D., Kobayashi A., Olsen A.O.,
RA   Carrano A.V.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Orphan receptor.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; AF024690; AAB86713.1; -.
DR   EMBL; AC002511; AAB67886.1; -.
DR   PIR; JC5717; JC5717.
DR   Genew; HGNC:4501; GPR43.
DR   MIM; 603823; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   GO; GO:0004930; F:G-protein coupled receptor activity; TAS.
DR   GO; GO:0007186; P:G-protein coupled receptor protein signalin. .; TAS.
```

```
RA   Sawzdargo M., George S.R., Nguyen T., Xu S., Kolakowski L.F. Jr.,
RA   O'Dowd B.F.;
RT   "A cluster of four novel human G protein-coupled receptor genes
RT   occurring in close proximity to CD22 gene on chromosome 19q13.1.";
RL   Biochem. Biophys. Res. Commun. 239:543-547(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: Orphan receptor.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; AF024688; AAB86711.1; -.
DR   EMBL; BC035657; AAH35657.1; -.
DR   PIR; JC5715; JC5715.
DR   Genew; HGNC:4499; GPR41.
DR   MIM; 603821; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   GO; GO:0004930; F:G-protein coupled receptor activity; TAS.
DR   GO; GO:0007186; P:G-protein coupled receptor protein signalin. .; TAS.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane.
FT   DOMAIN        1     19   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     20     40   1 (POTENTIAL).
FT   DOMAIN       41     47   CYTOPLASMIC.
FT   TRANSMEM     48     68   2 (POTENTIAL).
FT   DOMAIN       69     90   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     91    111   3 (POTENTIAL).
FT   DOMAIN      112    132   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    133    153   4 (POTENTIAL).
FT   DOMAIN      154    178   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    179    199   5 (POTENTIAL).
FT   DOMAIN      200    222   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    223    243   6 (POTENTIAL).
FT   DOMAIN      244    258   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    259    279   7 (POTENTIAL).
FT   DOMAIN      280    346   CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    166    166   N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   346 AA;  38649 MW;  B3B19D62D11B6BA1 CRC64;
```

```
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; TransmemBrane; Polymorphism.
FT   DOMAIN        1     12       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     13     33       1 (POTENTIAL).
FT   DOMAIN       34     41       CYTOPLASMIC.
FT   TRANSMEM     42     62       2 (POTENTIAL).
FT   DOMAIN       63     84       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     85    105       3 (POTENTIAL).
FT   DOMAIN      106    126       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    127    147       4 (POTENTIAL).
FT   DOMAIN      148    173       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    174    194       5 (POTENTIAL).
FT   DOMAIN      195    219       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    220    240       6 (POTENTIAL).
FT   DOMAIN      241    255       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    256    276       7 (POTENTIAL).
FT   DOMAIN      277    330       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    151    151       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    167    167       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARIANT     211    211       L -> H (in dbSNP:409093).
FT                                /FTId=VAR_011861.
SQ   SEQUENCE   330 AA;  37143 MW;  F4A8AC6AFBDF1E90 CRC64;

Query Match          21.0%;  Score 332; DB 1;  Length 330;
Best Local Similarity 31.0%;  Pred. No. 7.4e-18;
Matches  53; Conservative  93; Mismatches 120; Indels  34; Gaps  12;

Qy    12 YVAAFALGFPLNVLAIRGATAHARL-RLTPSLVYALNLGCSDLLLTVSLPLKAVEALASG 70
            ||:| |:|::|         ||:|
Db    14 YIIIFLTGLPANLLALRAFVGRIRQPQPAPVHILLSLTLADLLLLLLPFKIIEAASNF 73

Qy    71 AWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRPCYSWGVCAAI 130
Db    74 RWYLPKVVCALTSFGFYSSIYCSTWLLAGISIERYLGVAFPVQYKLSRRPLY--GVIAAL 131

Qy   131 --WALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPV-CLEAWDPASAG--PARF 184
Db   132 VAVWMSFGHCTIVLIVQ------YLNTTEQVR---SGNEITCYENFTDNQLDVVLPVRL 181

Qy   185 SLSLLLFFLPLAITAFCY--VGCLRALARSGLTHRRKLRAAWVAGGALLTLLLCVGPYN 241
Db   182 ELCIVLFFIPMAVTIFCYWRFVWIMLSQPLVGAQRRR--RAVGLAVVTLNLFLVCFGPYN 239

Qy   242 ASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGYLGR------GPGLKTVCAARTQG 295
Db   240 VSHLVGY-HQRKSPWWRSIAVVFSSLNASLDPLLFYFSSSVVRRAFGRGLQVL---RNQG 295
```

```
CC   -!- FUNCTION: High affinity receptor for activated thrombin coupled to
CC       G proteins that stimulate phosphoinositide hydrolysis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- PTM: A proteolytic cleavage generates a new amino terminus that
CC       functions as a tethered ligand.
CC   -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC   -----------------------------------------------------------------------
```

```
DR   EMBL; U09632; AAA18498.1; -.
DR   PIR; I51667; I51667.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; TransmemBrane; Glycoprotein; Signal;
KW   Blood coagulation.
FT   SIGNAL        1     20       POTENTIAL.
FT   PROPEP       21     42       REMOVED FOR RECEPTOR ACTIVATION.
FT   CHAIN        43    420       PROTEINASE ACTIVATED RECEPTOR 1.
FT   DOMAIN       43    101       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    102    127       1 (POTENTIAL).
FT   DOMAIN      128    136       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    137    156       2 (POTENTIAL).
FT   DOMAIN      157    175       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    176    197       3 (POTENTIAL).
FT   DOMAIN      198    217       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    218    238       4 (POTENTIAL).
FT   DOMAIN      239    267       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    268    287       5 (POTENTIAL).
FT   DOMAIN      288    310       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    311    333       6 (POTENTIAL).
FT   DOMAIN      334    345       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    346    369       7 (POTENTIAL).
FT   DOMAIN      370    420       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      295    300       POLY-SER.
FT   SITE         42     43       CLEAVAGE (BY THROMBIN).
FT   DISULFID    174    253       BY SIMILARITY.
FT   CARBOHYD     38     38       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     86     86       N-LINKED (GLCNAC...) (POTENTIAL).
FT   MUTAGEN      42     42       K->A: LACK OF ACTIVATION BY THROMBIN.
SQ   SEQUENCE   420 AA;  47435 MW;  D5163F56AFE12372 CRC64;

Query Match          15.4%;  Score 244; DB 1;  Length 420;
Best Local Similarity 25.0%;  Pred. No. 3.1e-11;
Matches  72; Conservative  44; Mismatches 126; Indels  46; Gaps  7;

Qy    11 LYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLPLKAVEALASG 70
Db   107 LYTVFIVGLPLNLLAIIIFLFKMKVR-KPAVVVMLNLAIADVFFVSVLPFKIAYHLSGN 165

Qy    71 AWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRPCYSWGVCAAI 130
Db   166 DWLFGPGMCRIVTAIFYCNMYCSVLLIASISVDRFLAVVYPMHSLSWRTMSRAYMACSFI 225

Qy   131 WALVLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSP-----VCLEAWDPASAGPARF 184
Db   226 WLISIA--------STIPLLVTEQTQKIPRLDITTCHDVLDLKDLKDFYIY 268

Qy   185 --SLSLLLFFLPLAITAFCYVGCLRALARSGLTHR-RKLRAAWVAGGALLTLLLCVGPY 240
Db   269 YFSSFCLLFFVPFITTICYIGIIRSLSSSSIENSCKKTRALFLAVVLCVFIICFGPT 328

Qy   241 N--------ASNVASFLYPNLGGSWRKLGLITGAWSVVLNPLVTGY 278
Db   329 NVLFLTHYLQEANEPLYFAY-------ILSACVGSVSCCLDPLIYYY 368
```

```
RESULT 5
PAR1_XENLA
ID   PAR1_XENLA     STANDARD;     PRT;    420 AA.
AC   P47749;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Proteinase activated receptor 1 precursor (PAR-1) (Thrombin receptor).
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94195429; PubMed=8145852;
RA   Gerszten R.E., Chen J., Ishii M., Ishii K., Nanevicz T.,
RA   Turck C.W., Vu T.-K.H., Coughlin S.R.;
RT   "Specificity of the thrombin receptor for agonist peptide is defined
RT   by its extracellular surface.";
RL   Nature 368:648-651(1994).
```

```
RESULT 6
PAR4_HUMAN
ID   PAR4_HUMAN    STANDARD;    PRT;   385 AA.
AC   Q96RI0; O76067;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Proteinase activated receptor 4 precursor (PAR-4) (Thrombin receptor-
DE   like 3) (Coagulation factor II receptor-like 3).
GN   F2RL3 OR PAR4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98389762; PubMed=9722561;
RA   Kahn M.L., Hammes S.R., Botka C., Coughlin S.R.;
RT   "Gene and locus structure and chromosomal localization of the
RT   protease-activated receptor gene family.";
RL   J. Biol. Chem. 273:23290-23296(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98379991; PubMed=9716134;
RA   Kahn M.L., Zheng Y.-W., Huang W., Bigornia V., Zeng D., Moff S.,
RA   Farese R.V. Jr., Tam C., Coughlin S.R.;
RT   "A dual thrombin receptor system for platelet activation.";
RL   Nature 394:690-694(1998).
RN   [3]
RP   SEQUENCE FROM N.A., AND MUTAGENESIS OF ARG-47 AND ARG-68.
RC   TISSUE=Lymphoma;
RX   MEDLINE=98283984; PubMed=9618465;
RA   Xu W.-F., Andersen H., Whitmore T.E., Presnell S.R., Yee D.P.,
RA   Ching A., Gilbert T., Davie E.W., Foster D.C.;
RT   "Cloning and characterization of human protease-activated receptor
RT   4.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:6642-6646(1998).
RN   [4]
RP   SEQUENCE FROM N.A., AND VARIANTS VAL-296 AND LEU-310.
RA   Rieder M.J., Carrington D.P., Chung M.-W., Lee K.L., Poel C.L., Yi Q.,
RA   Nickerson D.A.;
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   FUNCTION.
RX   MEDLINE=99178892; PubMed=10079109;
RA   Kahn M.L., Nakanishi-Matsui M., Shapiro M.J., Ishihara H.,
RA   Coughlin S.R.;
RT   "Protease-activated receptors 1 and 4 mediate activation of human
RT   platelets by thrombin.";
RL   J. Clin. Invest. 103:879-887(1999).
CC   -!- FUNCTION: Receptor for activated thrombin or trypsin coupled to G
CC       proteins that stimulate phosphoinositide hydrolysis. May play a
CC       role in platelets activation.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: Widely expressed, with highest levels in lung,
CC       pancreas, thyroid, testis and small intestine. Not expressed in
CC       brain, kidney, spinal cord and peripheral blood leukocytes. Also
CC       detected in platelets.
CC   -!- PTM: A proteolytic cleavage generates a new amino terminus that
CC       functions as a tethered ligand.
CC   -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF080214; AAC28860.1; -.
DR   EMBL; AF055917; AAC25699.1; -.
DR   EMBL; AF384819; AAK61908.1; -.
DR   Genew; HGNC:3540; F2RL3.
DR   MIM; 602779; -.
DR   InterPro; IPR000276; GPCR_Rhodopsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW   Blood coagulation; Polymorphism.
FT   SIGNAL        1     17       POTENTIAL.
FT   PROPEP       18     47       REMOVED FOR RECEPTOR ACTIVATION
FT                                (BY SIMILARITY).
FT   CHAIN        48    385       PROTEINASE ACTIVATED RECEPTOR 4.
FT   DOMAIN       48     82       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     83    103       1 (POTENTIAL).
FT   DOMAIN      104    108       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    109    129       2 (POTENTIAL).
FT   DOMAIN      130    151       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    152    172       3 (POTENTIAL).
FT   DOMAIN      173    192       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      193    213       4 (POTENTIAL).
FT   TRANSMEM    214    247       EXTRACELLULAR (POTENTIAL).
FT   DOMAIN      248    268       5 (POTENTIAL).
FT   DOMAIN      269    283       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    284    304       6 (POTENTIAL).
FT   DOMAIN      305    319       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    320    343       7 (POTENTIAL).
FT   DOMAIN      344    385       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      301    305       POLY-LEU.
FT   SITE         47     48       CLEAVAGE (BY THROMBIN OR TRYPSIN) (BY
FT                                SIMILARITY).
FT   DISULFID    149    228       BY SIMILARITY.
FT   CARBOHYD     56     56       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     296    296       F -> V.
FT                                /FTId=VAR_012852.
FT   VARIANT     310    310       P -> L.
FT                                /FTId=VAR_012853.
FT   MUTAGEN      47     47       R->A: NO PROTEOLYTIC CLEAVAGE (BY
FT                                THROMBIN OR TRYPSIN).
FT   MUTAGEN      68     68       R->A: NO EFFECT ON RECEPTOR ACTIVATION.
FT   CONFLICT    120    120       T -> A (IN REF. 4).
SQ   SEQUENCE   385 AA;  41162 MW;  FACA62B40D5756AE  CRC64;

Query Match             14.6%;  Score 231.5;  DB 1;  Length 385;
Best Local Similarity   24.9%;  Pred. No. 2.4e-10;
Matches    88;  Conservative   29;  Mismatches  123;  Indels  113;  Gaps    8;

Qy      3 LPPQLSFGLYVAAFALGFPLNVLAIRG-ATAHARLRLTPSLVYALNLGCSDLLLTVSLPL 61
          :|::|   |.|   |:|      ||   | :|||         ||::| :||    ::||
Db     75 VPTRLVPALYGLVLVVGLPANGLALWVLATQAPRL---PSTMLLMNLATADLLLALALPP 131

Qy     62 KAVEALASGAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRPC 121
          :  |  ::    ::|.         ||||          |||| ||| :|
Db    132 RIAYHLRGQRWPFGEAACRLATAALYGHMYGSVLLAAVSLDRYLALVHPLRARALRGRR 191

Qy    122 YSWGVCAAIWAL------------------------------VLCHGLVFGLEAPGGWLDHSNTSLG 159
          : |||: :                                  |||| :
Db    192 LALGLCMAAWLMAAALALPLTLQRQTFRLARSDRVLCHDALPLDAQA------------ 238

Qy    160 INTPVNGSPVCLEAWDPASAGPARFSLSLLFFLPLAITAFCYVGCLRALARSGLTHRRK 219
                                      :      |||:|
Db    239 ------------------SHWQPAFT----CLALLGCFLPLLAMLLCYGATIHTLAASGRRYGHA 281

Qy    220 LR--------------------------AAWVAGGALLTLLCVGPYNAS 243
          ||                          :|:
Db    282 LRLTAVVLASGAVAFFVPSNLLLLLHYSDPSPSAWGNLYGAVVPSLALSTLNSCVDP---- 337

Qy    244 NVASFLYPNLGGSWRKLGLITGAWSVLNPLVTGYLGRGPGLKTVCAARTCGG 296
          |:  | ::   ::|
Db    338 ----FIYYVSAEFR------------------------DKVRAGLFQRSPGDTVASKASAEGG 373
```

```
RESULT 7
PAR1_HUMAN
ID   PAR1_HUMAN     STANDARD;      PRT;    425 AA.
AC   P25116; Q96RF7; Q9BUN4;
DT   01-MAY-1992 (Rel. 22, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Proteinase activated receptor 1 precursor (PAR-1) (Thrombin receptor)
DE   (Coagulation factor II receptor).
GN   F2R OR PAR1 OR TR OR CF2R.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91168254; PubMed=1672265;
RA   Vu T.-K.H., Hung D.T., Wheaton V.I., Coughlin S.R.;
RT   "Molecular cloning of a functional thrombin receptor reveals a novel
RT   proteolytic mechanism of receptor activation.";
RL   Cell 64:1057-1068(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Rieder M.J., Carrington D.P., Chung M.-W., Lee K.L., Poel C.L., Yi Q.,
RA   Nickerson D.A.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Eye;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [4]
RP   FUNCTION.
RX   MEDLINE=99178892; PubMed=10079109;
RA   Kahn M.L., Nakanishi-Matsui M., Shapiro M.J., Ishihara H.,
RA   Coughlin S.R.;
RT   "Protease-activated receptors 1 and 4 mediate activation of human
RT   platelets by thrombin.";
RL   J. Clin. Invest. 103:879-887(1999).
RN   [5]
RP   VARIANT GLY-166.
RX   MEDLINE=99318093; PubMed=10391209;
RA   Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,
RA   Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nemesh J., Ziaugra N.,
RA   Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.Q.,
RA   Lander E.S.;
RT   "Characterization of single-nucleotide polymorphisms in coding regions
RT   of human genes.";
RL   Nat. Genet. 22:231-238(1999).
RN   [6]
RP   ERRATUM.
RA   Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,
RA   Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nemesh J., Ziaugra L.,
RA   Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.Q.,
RA   Lander E.S.;
RL   Nat. Genet. 23:373-373(1999).
CC   -!- FUNCTION: High affinity receptor for activated thrombin coupled to
CC       G proteins that stimulate phosphoinositide hydrolysis. May play a
CC       role in platelets activation and in vascular development.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: Platelets and vascular endothelial cells.
CC   -!- PTM: A proteolytic cleavage generates a new amino terminus that
CC       functions as a tethered ligand.
CC   -!- PTM: Phosphorylated; probably mediating desensitisation prior to
CC       the uncoupling and internalization of the receptor.
CC   -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M62424; AAA36743.1; -.
DR   EMBL; AF391809; AAK69768.1; -.
DR   EMBL; BC002464; AAH02464.1; -.
DR   EMBL; BC051909; AAH51909.1; -.
DR   PIR; A37912; A37912.
DR   PDB; 1NRN; 31-MAY-94.
DR   PDB; 1NRQ; 31-MAY-94.
DR   PDB; 1NRR; 31-MAY-94.
DR   Genew; HGNC:3537; F2R.
DR   MIM; 187930; -.
DR   GO; GO:0005794; C:Golgi apparatus; TAS.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   GO; GO:0005102; F:receptor binding; TAS.
DR   GO; GO:0015057; F:thrombin receptor activity; TAS.
DR   GO; GO:0006915; P:apoptosis; TAS.
DR   GO; GO:0007596; P:blood coagulation; TAS.
DR   GO; GO:0006919; P:caspase activation; TAS.
DR   GO; GO:0006928; P:cell motility; TAS.
DR   GO; GO:0007275; P:development; TAS.
DR   GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.
DR   GO; GO:0007125; P:invasive growth; TAS.
DR   GO; GO:0000074; P:regulation of cell cycle; TAS.
DR   GO; GO:0009611; P:response to wounding; TAS.
DR   GO; GO:0007165; P:signal transduction; TAS.
DR   GO; GO:0007262; P:STAT protein nuclear translocation; TAS.
DR   GO; GO:0007260; P:tyrosine phosphorylation of STAT protein; TAS.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW   Blood coagulation; Phosphorylation; Polymorphism; 3D-structure.
FT   SIGNAL        1     26       POTENTIAL.
FT   PROPEP       27     41       REMOVED FOR RECEPTOR ACTIVATION.
FT   CHAIN        42    425       PROTEINASE ACTIVATED RECEPTOR 1.
FT   DOMAIN       42    102       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    103    128       1 (POTENTIAL).
FT   DOMAIN      129    137       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      138    157       2 (POTENTIAL).
FT   TRANSMEM    158    176       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    177    198       3 (POTENTIAL).
FT   DOMAIN      199    218       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      219    239       4 (POTENTIAL).
FT   TRANSMEM    240    268       EXTRACELLULAR (POTENTIAL).
FT   DOMAIN      269    288       5 (POTENTIAL).
FT   TRANSMEM    289    311       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      312    334       6 (POTENTIAL).
FT   TRANSMEM    335    350       EXTRACELLULAR (POTENTIAL).
FT   DOMAIN      351    374       7 (POTENTIAL).
FT   DOMAIN      375    425       CYTOPLASMIC (POTENTIAL).
```

```
FT   DOMAIN      57     60      ASP/GLU-RICH (ACIDIC).
FT   SITE        41     42      CLEAVAGE (BY THROMBIN).
FT   DISULFID   175    254      BY SIMILARITY.
FT   CARBOHYD    35     35      N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    62     62      N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    75     75      N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD   250    250      N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD   259    259      N-LINKED (GLCNAC...)  (POTENTIAL).
FT   VARIANT    166    166      S -> G (IN REF. 1).
FT                              /FTId=VAR 014167.
FT   CONFLICT   238    238      L -> V (IN REF. 1).
FT   CONFLICT   335    335      A -> V (IN REF. 3; AAH02464).
FT   CONFLICT   364    364      C -> S (IN REF. 1).
SQ   SEQUENCE   425 AA;   47440 MW;  41B742A99EEC96AB CRC64;

Query Match             14.4%;  Score 228;  DB 1;  Length 425;
Best Local Similarity   25.3%;  Pred. No. 4.8e-10;
Matches   65;  Conservative  46;  Mismatches 104;  Indels  42;  Gaps  8;

Qy    11 LYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLPLKAVEALASG 70
            :|:::|||:||| | | :| :::|| | :||:::||||
Db   108 VYTGVFVVSLPVNIMAIVVFILKMKVK-KPAVVYMLHLATADVLFVSVLPFKISYYLSGS 166

Qy    71 AWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRPCYSWGVCAAI 130
            :||| :| | || : || |:: :| || :::||
Db   167 DWQFGSELCRFVTAAFYCNMYASILLMTVISIDRFLAVVYPMQSLSWRTLGRASFTCLAI 226

Qy   131 WALVLCHLGLV------FGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPA------- 177
            ||| :::  |       ::|||
Db   227 WALAIA--GVVPLLKEQTIQVPG----------LNITT------CHDVLNETLLEGYYA 268

Qy   178 ---SAGPARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHR-RKLRAAWVAGGALLTL 233
           ||:||:|| :::  |   | || :||
Db   269 YYFSA------FSAVFFVPLIISTVCYVSIIRCLSSSAVANRSKKSRALFLSAAVFCIF 322

Qy   234 LLCVGPYNASNVASFLY 250
            ::|||  :|| |:
Db   323 IICFGPTNILLIAHYSF 339
```

RESULT 8
PAR1_PAPHA
```
ID   PAR1_PAPHA    STANDARD;    PRT;   425 AA.
AC   P56488;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Proteinase activated receptor 1 precursor (PAR-1) (Thrombin receptor).
GN   P2R OR PAR1 OR BTHR12.
OS   Papio hamadryas (Hamadryas baboon).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Papio.
OX   NCBI_TaxID=9557;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Shoji M., Hayzer D.J., Hanson S.R.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
-!-  FUNCTION: High affinity receptor for activated thrombin coupled to
-!-  G proteins that stimulate phosphoinositide hydrolysis.
-!-  SUBCELLULAR LOCATION: Integral membrane protein.
-!-  PTM: A proteolytic cleavage generates a new amino terminus that
-!-  functions as a tethered ligand.
-!-  SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
```

```
DR   EMBL; AF028727; AAB84191.1; -.
DR   HSSP; P34996; 1DDD.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW   Blood coagulation.
FT   SIGNAL       1     26      POTENTIAL.
FT   PROPEP      27     41      REMOVED FOR RECEPTOR ACTIVATION.
FT   CHAIN       42    425      PROTEINASE ACTIVATED RECEPTOR 1.
FT   DOMAIN      42    102      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   103    128      1 (POTENTIAL).
FT   DOMAIN     129    137      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   138    157      2 (POTENTIAL).
FT   DOMAIN     158    176      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   177    198      3 (POTENTIAL).
FT   DOMAIN     199    218      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   219    239      4 (POTENTIAL).
FT   DOMAIN     240    268      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   269    288      5 (POTENTIAL).
FT   DOMAIN     289    311      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   312    334      6 (POTENTIAL).
FT   DOMAIN     335    350      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   351    374      7 (POTENTIAL).
FT   DOMAIN     375    425      CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      57     60      ASP/GLU-RICH (ACIDIC).
FT   SITE        41     42      CLEAVAGE (BY THROMBIN).
FT   DISULFID   175    254      BY SIMILARITY.
FT   CARBOHYD    35     35      N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    62     62      N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    75     75      N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD   250    250      N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD   259    259      N-LINKED (GLCNAC...)  (POTENTIAL).
SQ   SEQUENCE   425 AA;   47253 MW;  118FC5FB39D4DE2C CRC64;

Query Match             14.3%;  Score 226;  DB 1;  Length 425;
Best Local Similarity   25.3%;  Pred. No. 6.7e-10;
Matches   65;  Conservative  47;  Mismatches 103;  Indels  42;  Gaps  8;

Qy    11 LYVAARFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLPLKAVEALASG 70
            :|:::|||:||| | | :| :::|| | :||:::||||
Db   108 VYTGVFVVSLPVNIMAIVVFILKMKVK-KPAVVYMLHLATADVLFVSVLPFKISYYLSGS 166

Qy    71 AWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRPCYSWGVCAAI 130
            :||| :| | || : || |:: :| || :::||
Db   167 DWQFGSELCRFVTAAFYCNMYASILLMTVISIDRFLAVVYPMQSLSWRTLGRASFTCLAI 226

Qy   131 WALVLCHLGLV------FGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPA------- 177
            ||| :::  |       ::|||
Db   227 WALAIA--GVVPLLKEQTIQVPG----------LNITT------CHDVLNETLLEGYYA 268

Qy   178 ---SAGPARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHR-RKLRAAWVAGGALLTL 233
           ||:||:|| :::  |   | || :||
Db   269 YYFSA------FSAVFFVPLIISTVCYVSIIRCLSSSTVANRSKKSRALFLSAAVFCIF 322

Qy   234 LLCVGPYNASNVASFLY 250
            ::|||  :|| |:
Db   323 IICFGPTNILLIAHYSF 339
```

RESULT 9
GP92_HUMAN
```
ID   GP92_HUMAN    STANDARD;    PRT;   372 AA.
AC   Q9HICO;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Probable G protein-coupled receptor GPR92.
GN   GPR92 OR GPR93.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20517346; PubMed=11062477;
RA   White K.E., Evans W.E., O'Riordan J.L.H., Speer M.C., Econs M.J.,
RA   Lorenz-Depiereux B., Grabowski M., Meitinger T., Strom T.M.;
RT   "Autosomal dominant hypophosphataemic rickets is associated with
RT   mutations in FGF23.";
RL   Nat. Genet. 26:345-348(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21458557; PubMed=11574155;
RA   Lee D.K., Nguyen T., Lynch K.R., Cheng R., Vanti W.B., Arkhitko O.,
RA   Lewis T., Evans J.F., George S.R., O'Dowd B.F.;
RT   "Discovery and mapping of ten novel G protein-coupled receptor
RT   genes.";
RL   Gene 275:83-91(2001).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22040266; PubMed=12044878;
RA   Takeda S., Kadowaki S., Haga T., Takaesu H., Mitaku S.;
RT   "Identification of G protein-coupled receptor genes from the human
RT   genome sequence.";
RL   FEBS Lett. 520:97-101(2002).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=B-cell;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
-!- FUNCTION: Orphan receptor.
-!- SUBCELLULAR LOCATION: Integral membrane protein.
-!- TISSUE SPECIFICITY: Not expressed in frontal cortex, basal
CC      forebrain, caudate putamen, thalamus, or hippocampus.
-!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
-------------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
-------------------------------------------------------------------------------
DR   EMBL; AJ272207; CAC03715.1; -.
DR   EMBL; AF411112; AAL26483.1; -.
DR   EMBL; AB083600; BAB89313.1; -.
DR   EMBL; BC033571; AAH33571.1; -.
DR   Genew; HGNC:13307; GPR92.
DR   MIM; 606926; -.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
```

```
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein.
FT   DOMAIN        1     26       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     27     47       1 (POTENTIAL).
FT   DOMAIN       48     55       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     56     76       2 (POTENTIAL).
FT   DOMAIN       77     96       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     97    117       3 (POTENTIAL).
FT   DOMAIN      118    136       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    137    157       4 (POTENTIAL).
FT   DOMAIN      158    187       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    188    208       5 (POTENTIAL).
FT   DOMAIN      209    239       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    240    260       6 (POTENTIAL).
FT   DOMAIN      261    276       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    277    297       7 (POTENTIAL).
FT   DOMAIN      298    372       CYTOPLASMIC (POTENTIAL).
FT   DISULFID     94    175       BY SIMILARITY.
FT   CARBOHYD      4      4       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      9      9       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   372 AA;  41346 MW;  BA35709084BB6D84 CRC64;
```

```
Query Match             13.7%;  Score 217.5;  DB 1;  Length 372;
Best Local Similarity   29.1%;  Pred. No. 2.6e-09;
Matches   92;  Conservative   36;  Mismatches  141;  Indels   47;  Gaps   14;
```

```
Qy        6  QLSFGLYVAAFALGFPLNVLA----IRGATAHARLRLTPSLVYALNLGCSDLLLTVSLPL   61
             :|:|| ||:|||:|||||:|:    |:||  ||:| |:||:|:.||||||:|:|:|:
Db       23  RLHLVVYSLVLAAGLPLNALALWVFLRALRVHSVVS----VYMCNLAASDLLFTLSLPV   77

Qy       62  KAVEALASGAWPLEPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRPC  121
              :.:| :  :||||||| :  :    :  |    |.|| :|:  ||:|||:||:||
Db       78  R-LSYYALHHWPFFDLLCQTTGAIFQMNMYGSCIFMLINVDRYAAIVHPLRLRHLRRPR  136

Qy      122  YSWGVCAATWALVLCHLGLVFGLRAPGGWLDHSNTSLGINTPVNGSPVCLEAW-DPASAG  180
                |:||:  :   |:|::| |            |::|  :|| :  |   |  |::
Db      137  VARLLCLGVWALI----LVFAV--PAARVHRPSRCRYRDLEVR---LCFSFSDELWKG  186

Qy      181  --PARFSLSLLFFLPLAITAFCVYGCLRALARSGLT--HRRKLRAAWVAGGALLTLL  234
               :|  |:|||| :| :|:|: |:|:       |:|.  |||:|:|||:|||||||:
Db      187  RLLPLVLLABALGFLLPLAAVVYSSGRVFWTLARPDATQSQRRKTVRLLLAN--LVIFL  244

Qy      235  LCVGPYNA--------SNVASFLYP---NLGGSWRKLGLITGAWSVVLNPLVTGYLGRG  282
             || ||| :        :  :|  :|.    :||| | | :.||:::||||||| |::|
Db      245  LCFVYNSTLAVYGLLRSKLVAASVPARDRVRGVLMVVLLAGA-NCVLDPLVYFSAEG  303

Qy      283  -----PGLKTVCAART  293
                  ::|
Db      304  FRNTLRGLGTPHRART  319
```

```
RESULT 10
GP20_HUMAN
ID   GP20_HUMAN    STANDARD;    PRT;    358 AA.
AC   Q99678;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Probable G protein-coupled receptor GPR20.
GN   GPR20.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97225799; PubMed=9073069;
RA   O'Dowd B.F., Nguyen T., Jung B.P., Marchese A., Cheng R.,
RA   Heng H.H.Q., Kolakowski L.F. Jr., Lynch K.R., George S.R.;
RT   "Cloning and chromosomal mapping of four putative novel human
RT   G-protein-coupled receptor genes.";
RL   Gene 187:75-81(1997).
CC   -!- FUNCTION: Orphan receptor.
```

```
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: LIVER AND THE BRAIN REGIONS THALAMUS, PUTAMEN,
CC       AND CAUDATE, BUT NOT IN FRONTAL CORTEX, PONS AND HYPOTHALAMUS.
CC   -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; U66579; AAC51302.1; -.
DR   Genew; HGNC:4475; GPR20.
DR   MIM; 601908; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   GO; GO:0004930; F:G-protein coupled receptor activity; TAS.
DR   GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; TAS.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; TransmemBrane; Glycoprotein.
FT   DOMAIN        1     48       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     49     69       1 (POTENTIAL).
FT   DOMAIN       70     86       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     87    107       2 (POTENTIAL).
FT   DOMAIN      108    125       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    126    146       3 (POTENTIAL).
FT   DOMAIN      147    168       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    169    189       4 (POTENTIAL).
FT   DOMAIN      190    196       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    197    217       5 (POTENTIAL).
FT   DOMAIN      218    238       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    239    259       6 (POTENTIAL).
FT   DOMAIN      260    275       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    276    296       7 (POTENTIAL).
FT   DOMAIN      297    358       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD     16     16       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     26     26       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   358 AA;  38406 MW;  258C6683D455B1EB CRC64;

Query Match           13.7%;  Score 216.5; DB 1; Length 358;
Best Local Similarity 29.0%;  Pred. No. 2.9e-09;
Matches   73; Conservative  37; Mismatches  87; Indels  55; Gaps    8;

Qy     14 AAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLPLK-AVEALASG-- 70
              ||  |||   :: ||||:|: :|||||| :|||.: :||| | |:: ||| |
Db     62 AIFLAGLVLNGLALYVFCCRTRAK-TPSVIYTINLVVTDLLVGLSLPTRFAVVYGARGCL 120

Qy     71 --AWPLPASLCPVPAVAHFPLYAGGGFLAALSAGRYLGAAFPLGYQAFRRPCYSWGVCA 128
               :||| :: |   |||||| || :|.||    .|:   || :: |:: ||| | ||
Db    121 RCAFP-------HVLGYFLNMHCSILFLTCICVDRYLAIVRPBAPAACRQPACARAVCA 172

Qy    129 AIW---ALVLCHGLGVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLRAWDPSAGPARF 184
               ||     |: ||  | :                                |    :
Db    173 FVWLAAGAVTLSVLG------------------------VTGSRPCCRV-------F 198

Qy    185 SLSLLLFLPLAITAFCYVGCLRALARSGLITH--RRKLRAAWVAGGALLTLLLCVGPYN 241
           : |:| |||| ||  |.:  |||| :|| :     | |:: :|| ||| | |::
Db    199 ALTVLEFLLPLLVISVFTGRIMCALSRPGLLHQGRQRRVRAMQLLLTVLIIFLVCFTPFH 258

Qy    242 ASNVASFLYPNL 253
                :|:|
Db    259 ARQVAVALWPDM 270

RESULT 11
PAR1_CRILO
ID   PAR1_CRILO    STANDARD;    PRT;   428 AA.
AC   Q00991; Q60461;
```

```
DT   01-APR-1993 (Rel. 25, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Proteinase activated receptor 1 precursor (PAR-1) (Thrombin receptor).
GN   F2R OR PAR1.
OS   Cricetulus longicaudatus (Long-tailed hamster) (Chinese hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Cricetulus.
OX   NCBI_TaxID=10030;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RX   MEDLINE=91348247; PubMed=1652467;
RA   Rasmussen U.B., Vouret-Craviari V., Jallat S., Schlesinger Y.,
RA   Pages G., Pavirani A., Lecocq J.P., Pouyssegur J.,
RA   Obberghen-Schilling B.;
RT   "cDNA cloning and expression of a hamster alpha-thrombin receptor
RT   coupled to Ca2+ mobilization.";
RL   FEBS Lett. 288:123-128(1991).
RN   [2]
RP   SEQUENCE OF 42-428 FROM N.A.
RX   MEDLINE=96028007; PubMed=7488069;
RA   Hartmann T., Grace M.B., Buzard G.S., Ruoss S.J.;
RT   "Thrombin receptor polymorphism in Chinese hamster.";
RL   Biochem. Biophys. Res. Commun. 215:974-980(1995).
CC   -!- FUNCTION: High affinity receptor for activated thrombin coupled to
CC       G proteins that stimulate phosphoinositide hydrolysis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- PTM: A proteolytic cleavage generates a new amino terminus that
CC       functions as a tethered ligand.
CC   -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; X61958; CAA43957.1; -.
DR   EMBL; U34047; AAA86747.1; -.
DR   PIR; S17148; S17148.
DR   HSSP; P34996; 1DDD.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; TransmemBrane; Glycoprotein; Signal;
KW   Blood coagulation; Polymorphism.
FT   SIGNAL        1     21       POTENTIAL.
FT   PROPEP       22     41       REMOVED FOR RECEPTOR ACTIVATION (BY
FT                                SIMILARITY).
FT   CHAIN        42    428       PROTEINASE ACTIVATED RECEPTOR 1.
FT   DOMAIN       42    105       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    106    131       1 (POTENTIAL).
FT   DOMAIN      132    140       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    141    160       2 (POTENTIAL).
FT   DOMAIN      161    179       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    180    201       3 (POTENTIAL).
FT   DOMAIN      202    221       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      222    242       4 (POTENTIAL).
FT   TRANSMEM    243    271       EXTRACELLULAR (POTENTIAL).
FT   DOMAIN      272    291       5 (POTENTIAL).
FT   DOMAIN      292    314       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    315    337       6 (POTENTIAL).
FT   DOMAIN      338    352       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    353    377       7 (POTENTIAL).
FT   DOMAIN      378    428       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       60     63       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      301    304       POLY-SER.
```

```
FT   SITE             41     42       CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT   DISULFID        178    257       BY SIMILARITY.
FT   CARBOHYD         65     65       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD         78     78       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD        253    253       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD        262    262       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARIANT          84     84       MISSING (IN AN ALLELE).
FT   CONFLICT        384    384       H -> T (IN REF. 1).
SQ   SEQUENCE        428 AA;  47602 MW;  E5492AB2CD877E2F CRC64;

Query Match             13.4%;  Score 213;  DB 1;  Length 428;
Best Local Similarity   23.4%;  Pred. No. 6.2e-09;
Matches   68;  Conservative   53;  Mismatches  122;  Indels  48;  Gaps   9;

Qy     11 LYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLPLKAVEALASG 70
          :|..||||:|..||:||:||
Db    111 VTTFVFVVSLPLNILAIAVFVLKMKVK-KPAVVYMLHLAMADVLFVSVLPLKISYYFSGS 169

Qy     71 AWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGVQAFRRPCYSWGVCAAI 130
          :.|::.||..:||   ..|:.||:.|:.|.||.::|.  |.:|
Db    170 DWQFGSGMCRFATAAFYCNMYASIMLMTVISIDRFLAVVYPIQSLSWRTLGRANFTCLVI 229

Qy    131 WALVLCHLGLVFGL------EAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPA------- 177
                                                    :.
Db    230 WVMAI--MGVVPLLLKEQTRVPG-------LNITT------CHDVLNETLLQGFVS 271

Qy    178 ---SAGPARFSLSLLLFFLPLAITAFCYVGCLRALARSGLTHR-RKLRAAWVAGGALLTL 233
             ||..|:  ||.:::||: .|:|..| .::
Db    272 YYPSA------FSAVFFLVPLIISTICVMSIIRCLSSSSVANRSKKSRALFLSAAVFCVF 325

Qy    234 LLCVGPYNASNVASFLYPNLGGSWRK------LGLITGAWSVVLNPLVTGY 278
          :||.|||| | |||  ::|  .||:|        ||:
Db    326 IVCFGPTNVLLIMHYLLLSDSPATEKAYFAYLLCVCVSSVSCCIDPLIYYY 376
```

```
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; FALSE_NEG.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein.
FT   DOMAIN            1     22       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM         23     43       1 (POTENTIAL).
FT   DOMAIN           44     57       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM         58     78       2 (POTENTIAL).
FT   DOMAIN           79     96       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM         97    117       3 (POTENTIAL).
FT   DOMAIN          118    139       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM        140    160       4 (POTENTIAL).
FT   DOMAIN          161    189       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM        190    210       5 (POTENTIAL).
FT   DOMAIN          211    231       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM        232    252       6 (POTENTIAL).
FT   DOMAIN          253    275       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM        276    298       7 (POTENTIAL).
FT   DOMAIN          299    323       EXTRACELLULAR (POTENTIAL).
FT   CARBOHYD          5      5       N-LINKED (GLCNAC...) (POTENTIAL).
FT   DISULFID         94    172       BY SIMILARITY.
SQ   SEQUENCE        328 AA;  37594 MW;  B74D49B99C7164A5 CRC64;

Query Match             13.1%;  Score 208;  DB 1;  Length 328;
Best Local Similarity   30.3%;  Pred. No. 1.2e-08;
Matches   67;  Conservative   25;  Mismatches  105;  Indels  24;  Gaps   7;

Qy     11 LYVAAFALGFPLNVLAIRGATAHARLRLTPSLVYALNLGCSDLLLTVSLPLKAVEALASG 70
          :|..||||:|..||:||:|||   .||:.||:.|
Db     27 VYSVVFLLGLPLNAVVI-GQIWLARKALRTTIYMLNLATADLLYVCSLPLLIYNYTQKD 85

Qy     71 AWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPL-GYQAFRRPCYSWGVCAA 129
          ...:| |.||..:|.| ..:||| ::|.||..:|.|
Db     86 YWPFGDFTCKFVRFQFYTNLHGSILFLTCISVQRTMGICHPLASWHKKKGKKLTWLVCAA 145

Qy    130 IWALVLCHLGLVFGLEAPGWLDHSNTSLGINTPVNGSPVCLEAWDP---ASAGPARFSL 186
          :|.:|.   :.   .:|
Db    146 VWFIVIAQCLPTFVFASTG-------TQRNRT-VCYDLSPPDRSASYFPYGITL 191

Qy    187 SLLLFLPLAITAFCYVGCLRALARS----GL-THRRKLRA 222
          :::|:|||..||.|||..: .||    |:.||||:|
Db    192 TITGFLLPFAAILACYCSMARILCQKDELIGLAVHKKKDKA 232
```

```
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N; TISSUE=Mammary gland;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [4]
RP   SEQUENCE OF 1-74 FROM N.A.
RC   STRAIN=129/Sv;
RX   MEDLINE=96327649; PubMed=8678993;
RA   Xue J., Jenkins N.A., Gilbert D.J., Copeland N.G., Sadler J.B.;
RT   "Structure and localization of the thrombin receptor gene on mouse
RT   chromosome 13.";
RL   Mamm. Genome 7:625-626(1996).
CC   -!- FUNCTION: High affinity receptor for activated thrombin coupled to
CC       G proteins that stimulate phosphoinositide hydrolysis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- PTM: A proteolytic cleavage generates a new amino terminus that
CC       functions as a tethered ligand.
CC   -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; L03529; AAA40438.1; -.
DR   EMBL; U36757; AAB38308.1; -.
DR   EMBL; U36756; AAB38308.1; JOINED.
DR   EMBL; BC031516; AAH31516.1; -.
DR   EMBL; U55076; AAB00198.1; -.
DR   EMBL; U55075; AAB00198.1; JOINED.
DR   HSSP; P34996; 1DDD.
DR   MGD; MGI:101802; P2r.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW   Blood coagulation.
FT   SIGNAL        1     21       POTENTIAL.
FT   PROPEP       22     41       REMOVED FOR RECEPTOR ACTIVATION (BY
FT                                SIMILARITY).
FT   CHAIN        42    430       PROTEINASE ACTIVATED RECEPTOR 1.
FT   DOMAIN       42    107       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    108    133       1 (POTENTIAL).
FT   DOMAIN      134    142       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    143    162       2 (POTENTIAL).
FT   DOMAIN      163    181       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    182    203       3 (POTENTIAL).
FT   DOMAIN      204    223       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    224    244       4 (POTENTIAL).
FT   DOMAIN      245    273       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    274    293       5 (POTENTIAL).
FT   DOMAIN      294    316       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      317    339       6 (POTENTIAL).
FT   DOMAIN      340    354       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    355    379       7 (POTENTIAL).
FT   DOMAIN      380    430       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       60     65       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN       88     91       POLY-PRO.
FT   SITE         41     42       CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT   DISULFID    180    259       BY SIMILARITY.
FT   CARBOHYD     67     67       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     80     80       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    255    255       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CONFLICT    162    162       F -> S (IN REF. 1).
FT   CONFLICT    189    189       G -> Y (IN REF. 1).
FT   CONFLICT    223    223       R -> D (IN REF. 1).
FT   CONFLICT    262    262       V -> L (IN REF. 1).
FT   CONFLICT    365    365       S -> T (IN REF. 1).
SQ   SEQUENCE   430 AA;  47790 MW;  395FD64FAB52C9BF CRC64;
```

Query Match   13.1%; Score 208; DB 1; Length 430;
Best Local Similarity   25.1%; Pred. No. 1.5e-08;
Matches   64; Conservative 45; Mismatches 108; Indels 38; Gaps   9;

```
Qy          11 LYVAAFALGFPLNVLAIRGATAHARLRL-TPSLVYALNLGCSDLLTVSLPLPKAVEALAS  69
                  :|::  ||||||||       |:::  |::||::  :::.::||::
Db         113 VYTIVPFIVSLPLNVLAI--AVFVLRMKVKPAVVVMLHLAWADVLFVSVLPFKISYFSG 170

Qy          70 GAWPLPASLCPVFAVAHFFPLYAGGGFLAALSAGRYLGAAFPLGYQAPRRPCYSWGVCAA 129
                :  :   |::|   ||:   :|||  | ||| : :|:  || ::|||||
Db         171 TDWQPFGSGMCRFATARAFYGNMYASIMLMTVISIDRFLAVVYPIQSLSWRTLGRANFTCVV 230

Qy         130 IWALVLCHLGLVFGL------EAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAGPAR 183
                ||::||:|:  |||:||           :  |:|||| |:|:|
Db         231 IWVMAI--MGVVPLLLKEQTRVPG------------LNITT-------CHDVLSENLMQ 269

Qy         184 ------FS-LSLLLFPLPLAITAFCYVGCLRALARSGLTHR-RKLRAAWVAGGALLTLLL 235
                      |:  :|||||||||||::    |::::|::|:||::|  ::|:|:::|: ||:|:|
Db         270 GFYSYFSAFSAIFFLVPLIVSTVCYTSIIRCLSSSAVANRSKKSRALPLSAAVFCIFIV 329

Qy         236 CVGPYNASNVASFLY 250
                ||: |  ::|: :|:
Db         236 CFGPYNVLLIVHYLF 344
```

```
RESULT 14
SSR5_RAT
ID   SSR5_RAT     STANDARD;     PRT;     363 AA.
AC   P30938;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Somatostatin receptor type 5 (SSS5R).
GN   SSTR5.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pituitary;
RX   MEDLINE=93125499; PubMed=1362243;
RA   O'Carroll A.-M., Lolait S.J., Konig M., Mahan L.C.;
RT   "Molecular cloning and expression of a pituitary somatostatin
RT   receptor with preferential affinity for somatostatin-28.";
RL   Mol. Pharmacol. 42:939-946(1992).
RN   [2]
RP   REVISIONS TO C-TERMINUS.
RC   TISSUE=Pituitary;
RX   MEDLINE=94195267; PubMed=7908405;
RA   Panetta R., Greenwood M.T., Warszynska A., Demchyshyn L.L., Day R.,
```

```
RA    Niznik H.B., Srikant C.B., Patel Y.C.;
RT    "Molecular cloning, functional characterization, and chromosomal
RT    localization of a human somatostatin receptor (somatostatin receptor
RT    type 5) with preferential affinity for somatostatin-28.";
RL    Mol. Pharmacol. 45:417-427(1994).
-!-   FUNCTION: Receptor for somatostatin-28. The activity of this
      receptor is mediated by G proteins which inhibit adenylyl cyclase.
-!-   SUBCELLULAR LOCATION: Integral membrane protein.
-!-   TISSUE SPECIFICITY: Prominent in the pituitary and small
      intestine. Low levels in islets and spleen. Not detected in
      kidney, pancreas, cerebellum, or cortex.
-!-   SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
--------------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC    the European Bioinformatics Institute.  There are no restrictions on  its
CC    use by non-profit institutions as long  as its content is in  no  way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
--------------------------------------------------------------------------
DR    EMBL; L04535; AAA17029.1; -.
DR    EMBL; U01152; AAC09011.1; -.
DR    EMBL; X74828; CAA52825.1; -.
DR    PIR; I57940; I57940.
DR    HSSP; P02699; 1F88.
DR    InterPro; IPR000276; GPCR_Rhodpsn.
DR    Pfam; PF00001; 7tm_1; 1.
DR    PRINTS; PR00237; GPCRRHODOPSN.
DR    PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR    PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW    G-protein coupled receptor; Transmembrane; Glycoprotein;
KW    Multigene family; Lipoprotein; Palmitate.
FT    DOMAIN        1     35       EXTRACELLULAR (POTENTIAL).
FT    TRANSMEM     36     63       1 (POTENTIAL).
FT    DOMAIN       64     73       CYTOPLASMIC (POTENTIAL).
FT    TRANSMEM     74     99       2 (POTENTIAL).
FT    DOMAIN      100    111       EXTRACELLULAR (POTENTIAL).
FT    TRANSMEM    112    133       3 (POTENTIAL).
FT    DOMAIN      134    155       CYTOPLASMIC (POTENTIAL).
FT    TRANSMEM    156    176       4 (POTENTIAL).
FT    DOMAIN      177    196       EXTRACELLULAR (POTENTIAL).
FT    TRANSMEM    197    221       5 (POTENTIAL).
FT    DOMAIN      222    247       CYTOPLASMIC (POTENTIAL).
FT    TRANSMEM    248    273       6 (POTENTIAL).
FT    DOMAIN      274    283       EXTRACELLULAR (POTENTIAL).
FT    TRANSMEM    284    308       7 (POTENTIAL).
FT    DOMAIN      309    363       CYTOPLASMIC (POTENTIAL).
FT    CARBOHYD     13     13       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT    CARBOHYD     23     23       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT    CARBOHYD    186    186       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT    DISULFID    110    185       BY SIMILARITY.
FT    LIPID       320    320       S-palmitoyl cysteine (Potential).
SQ    SEQUENCE   363 AA;  39971 MW;  4BD451296O613B4A CRC64;

Query Match             13.0%;  Score 205.5;  DB 1;  Length 363;
Best Local Similarity   25.4%;  Pred. No. 1.9e-08;
Matches   78; Conservative  49; Mismatches 135;  Indels  45;  Gaps  12;

Qy   11 LYVAAFAGFPLNVLAIRGATAHARLRLTPSLVVALNLGCSDLLTVSLPLKAVEALASG 70
         ||:      : | ||   ||||:::| | |||  ||   | :|
Db   43 LYLLVCTVGLSGNTLVIVYVLRHAKMK-TVTNVYILNLAVADVLFMLGLPFLATQNAVVS 101

Qy   71 AWPLPASLCPVF---AVAHFPPLYAGGGPLAALSAGRYLGAAPPLGYQAFRRPCYSWGV 126
          |    :         ::   :||   |:||       ||||
Db  102 YWPPGSFLCRLVWTLDGINQFTSIFC----LMVMSVDRYLAVVHPLRSARWRRPRVAKMA 157

Qy  127 CAAIWAL-VLCHLGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAW-DPASAGPARF 184
         || |     :|: |||:    || |       :        : |  :
Db  158 SAAVWVFSLLMSLPLLVFADVQEGW------------------GTCNLSWPEPVGLWGAAF 200

Qy  185 --SLSLLLFFLPLAITAFCYVGCLRALARSGL--THRRKLRAAWVAGGALLTLLCVG- 238
           :|| ||| | | :  | :::  |||     :   :  :  :|| ||| :|: ||
```

```
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein;
KW   Chemotaxis.
FT   DOMAIN          1     40     EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM        41    67     1 (POTENTIAL).
FT   DOMAIN          68    73     CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM        74    92     2 (POTENTIAL).
FT   DOMAIN          93    114    EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM        115   138    3 (POTENTIAL).
FT   DOMAIN          139   159    CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM        160   184    4 (POTENTIAL).
FT   DOMAIN          185   204    EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM        205   232    5 (POTENTIAL).
FT   DOMAIN          233   247    CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM        248   270    6 (POTENTIAL).
FT   DOMAIN          271   290    EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM        291   313    7 (POTENTIAL).
FT   DOMAIN          314   355    CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD        7     7      N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD        21    21     N-LINKED (GLCNAC...) (POTENTIAL).
FT   DISULFID        115   192    BY SIMILARITY.
FT   CONFLICT        90    111    DLLFALTMPIWAVSKEKGWIFG ->
FT                                PAFCPDHAYLGRLQGKRLDFR (IN REF. 2).
FT   CONFLICT        146   147    HA -> QS (IN REF. 2).
FT   CONFLICT        204   204    R -> C (IN REF. 2).
FT   CONFLICT        287   288    DI -> EL (IN REF. 2).
SQ   SEQUENCE   355 AA;  40622 MW;  EFE49ACB9D1E0F21 CRC64;
```

```
Query Match              12.9%;   Score 204.5;  DB 1;  Length 355;
Best Local Similarity    25.8%;   Pred. No. 2.3e-08;
Matches   83;  Conservative   39;  Mismatches 117;  Indels  83;  Gaps   12;

Qy     13  VAAPALGPPLNVLA---IRGATAHARLRLTPSLVYALNLGCSDLLLTVSLPLKAVEALAS  69
               ::||| |::|       :|     ::|||| ||| :||| :::|::||
Db     48  VVIYALVFLLSLLGNSLVMLVILYSRSNRSVTDVYLLNLAMADLLFALTMPIWAVS--KE 105

Qy     70  GAWPLPASLCPVFAVAHFPPLYAGGGFLAALSAGRYLGAAFPPLGYQAFRRPCYSWGVCAA 129
               :||||| :: ::|       |||| |||    |:  : :|
Db    106  KGWIFGTFPLCKVVSLVKEVNFYSGILLLACISVDRYLAIVHATRTLTQKRHLVKF-ICLG 164

Qy    130  IWALVLCHGLVFGLEAPGGWLDHSNTSLGINTPVNGSPVCLEAWDPASAGPARFSLSLL 189
           ||||||| |||||||          :|| ||||    |: |     |:  |
Db    165  IWALSLI-LSLPFFLFRQ------------VFSPNNSSPVC---YBDLGHNTAKWRMVLR 208

Qy    190  L------FFLPLAITAFCYVGCLRAL--ARSGLTHRRKLRAAWVAGGALLTLLLCVGPYN 241
               :       |||| ||| :::||     ||      ::     |||| ||| :|
Db    209  ILPHTFGFLPLLVMLFCYGFTLRTLFQAHMGQKH----RAMRVIFAVVLIFLLCWLPYN 264

Qy    242  ----------------------------------ASNVASFLYPNLGGSWRKLGLITGAWSVVLN 272
                                              ::  ||
Db    265  LVLLADTLMRTHVIQETCQRRNDIDRALDATEILGFLHS-------------CLN 306

Qy    273  PLVTGYLGRG--PGLKTVCAAR 292
           |:::::||:       :  |||
Db    307  PIIYAFIGQNFRNGFLKMLAAR 328
```

Search completed: April 29, 2004, 12:10:33
Job time : 19 secs

THIS PAGE BLANK (USPTO)